



UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN
SCIENZE BIOMEDICHE

CICLO XXVII

COORDINATORE Prof. Persio Dello Sbarba

*Aspetti genetici ed epigenetici dell'infertilità
maschile*

Settore Scientifico Disciplinare MED/13

Dottorando

Dott. Chianese Chiara

(firma)

Tutore

Prof. Csilla Krausz

(firma)

Coordinatore

Prof. Gianni Forti

(firma)

Anni 2012/2014

Ai miei genitori

Ringraziamenti

Sembra ieri che tutto questo ha avuto inizio, e invece già ci sono...eccomi al traguardo. Un traguardo che non avrei potuto raggiungere senza l'appoggio e l'affetto delle persone che ho incrociato per il cammino e alle quali va la mia più sentita gratitudine.

Desidero innanzitutto ringraziare la Prof. Csilla Krausz, che sin dall'inizio ha riposto in me la sua fiducia, spingendomi verso mete che neanch'io credevo di poter raggiungere. Grazie per aver creduto in me e per avermi seguito con affetto durante questa importantissima tappa della vita.

Ringrazio il Prof. Dello Sbarba e Giulia Pellegrini, che mi hanno assistito egregiamente nella gestione del mio soggiorno all'estero. Ringrazio il Prof. Forti, per la sua comprensione e disponibilità.

Ringrazio il gruppo di Firenze, che per buona parte di questi anni è stato per me rifugio nei momenti di debolezza e svago nei momenti di allegria.

Claudia, grazie per esserci ancora dopo tutti questi anni: della serie "il primo tutor non si scorda mai!". Serena, grazie per avermi dimostrato il tuo affetto ascoltandomi nei momenti critici...e ovviamente per le "*girathine pe' tutte le sedi della Pubblica assistenza di Praho*". Fabrice, grazie perché molto di quello che so, lo devo a te. Elenina e Francesca, grazie per la vostra disponibilità incondizionata e per aver condiviso con me i magici momenti di caccia alle cartelle! Elenona, grazie per portare sempre il buon umore nei corridoi grigi *de' i' ccubo*! Sarina, grazie per le grasse risate dentro e fuori dal lab. Antoni, gràcies per la teva actitud sempre positiva i per l'entusiasme que poses en tot allò que fas...i sobretot, gràcies perquè encara que em vas conèixer en castellà, sempre em contestes en català! Tonino, grazie per esserci sempre, anche solo per fare due chiacchiere. Grazie Monica per le nostre chiacchierate, scientifiche e non. Grazie Marisa per il tuo contributo costante. Grazie Erminio e Selene per alleviare gli spiriti in laboratorio. Grazie a tutte le Zie (Elena, Lara, Marta, Sara, Giulia, Ilaria) per aver condiviso i momenti più belli delle loro vite e avermi fatto sentire vicina anche a chilometri di distanza. Un ringraziamento speciale a Marta, la mia informatrice *in situ*: grazie per ogni volta che mi hai avvisato di scadenze e relazioni di cui io ero puntualmente ignara.

Grazie alla mia famiglia Fiorentina: Nerone, Francesca e Chiara, in questa tesi ci siete anche voi...

Gracias a las chicas del laboratorio en Barcelona por haberme acogido con afecto desde el primer día que entré a la Fundació Puigvert. Elisabet, gracias para

intentar dar siempre tu ayuda constructiva. Gemma, gracias por compartir ideas, científicas y non. Rosalia, Ana y Olga, gracias por cuidarme con cariño las muestras. Patricia, gracias por tus mojitos! Elena, gracias por escuchar, a veces es lo único que hace falta. Irene, gracias por transmitir tu entusiasmo y simpatía al sólo nombrarte. Olga, ha sido una suerte conocerte. Luz, gracias por ser mi rayo de luz diario. Ania, gracias por compartir tu ánimo dulce (y tus bolis!) conmigo. Bea, gracias por tu amistad y tu apoyo todas las veces que lo he necesitado.

In particolare, De-bo-ri-naaaaah...grazie per esserci stata al mio arrivo alla Puccioverde e avermi fatto sentire a casa in un paese che non è il mio...bello averti incotrato e averti reso parte della mia vita....un giorno *ce ne jamm' a mmar'*!

Ringrazio le mie amiche di sempre, Martina e Noemi, quelle che non si dimenticano mai e che non ti dimenticano mai.

Il ringraziamento più speciale va alla mia famiglia. L'amore per Roberta, per mia madre e per mio padre è il motore che da sempre mi spinge a fare il possibile affinché siano fieri di me. Grazie per avermi dato i mezzi e le possibilità per andare avanti; grazie per avermi insegnato i valori necessari per affrontare le situazioni della vita. A voi dedico questa tesi e il lavoro svolto nell'arco di questi anni, che non avrei potuto affrontare senza il vostro infallibile e incondizionato sostegno.

Grazie ad Astor, il cucciolo tanto atteso che ha riempito di gioia gli ultimi giorni di stesura di questa tesi.

Infine, ma più di tutti, ringrazio Fede, il centro del mio mondo. Grazie per stare al mio fianco e per avermi tirato su nei momenti di *bajón*. Grazie per aver portato colore alla mia vita e per aver dato un senso a questi ultimi anni. Grazie per avermi aperto gli occhi le volte che mi accecavo di negatività e per avermi insegnato a prendere le cose con più filosofia. Semplicemente grazie, perché sei la persona più bella che conosca, dalla quale ho imparato tanto e dalla quale ho ancora tanto da imparare. E chissà, un giorno imparerò a giocare anche a scacchi...

INDEX

1. SUMMARY	3
2. INTRODUCTION	6
2.1 Male Infertility	6
2.1.1. Definition and Etiology.....	6
2.1.2 Genetics of Male Infertility.....	7
2.1.3. Epigenetics Aspects	12
2.2. Copy Number Variations (CNVs).....	17
2.2.1. Discovery and Definition	17
2.2.2. What are CNVs?	18
2.2.2. Potential Mechanisms of CNV Generation	19
2.2.3. CNVs and Disease: Functional Consequences	24
2.3 The Y Chromosome	27
2.3.1. Structure and gene content	27
2.3.2. Y chromosome-linked CNVs.....	33
2.4. The X chromosome.....	40
2.4.1. General features and structure.....	40
2.4.2. X-linked genes	41
3. AIM OF THE THESIS	45
4. RESULTS	46
4.1. PAPER 1.....	48
<i>High-resolution X chromosome-specific array-CGH detects new CNVs in infertile males.</i>	<i>48</i>
4.2. PAPER 2.....	62
<i>Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility.....</i>	<i>62</i>
4.3. PAPER 3.....	68
<i>X chromosome-linked CNVs in male infertility: discovery of overall duplication load and recurrent, patient-specific gains with potential clinical relevance.</i>	<i>68</i>
4.4. PAPER 4.....	76
<i>Y-chromosome microdeletions are not associated with SHOX haploinsufficiency.</i>	<i>76</i>
4.5. PAPER 5.....	83

<i>Novel insights into DNA methylation features in spermatozoa: stability and peculiarities.....</i>	83
5. DISCUSSION	100
5.1. X-linked CNVs.....	100
5.2. Y-chromosome microdeletions and <i>SHOX</i> CNVs.....	106
5.3. The “normal” sperm methylome	108
6. CONCLUSIONS.....	111
7. BIBLIOGRAPHY	113
Annexes.....	129

1. SUMMARY

Male infertility affects 7% of the general population and the etiology is most frequently related to primary testicular failure. Half of the cases of impaired spermatogenesis still remain unexplained and are referred to as idiopathic male factor infertility; such cases are likely to conceal a not yet identified genetic or epigenetic origin. This thesis aims at studying the genetic and epigenetic aspects of spermatogenesis, with the ultimate scope of identifying novel causes potentially leading to male infertility.

From the genetic standpoint, the enrichment on the Y and X chromosomes of genes involved in gonadal development/differentiation and/or predominantly expressed in the testis made both sex chromosomes an especially promising target of investigation in the attempt of discovering new genetic factors leading to male infertility. In this thesis, special focus was given to Copy Number Variations (CNVs), which have been reported as important contributors to a number of complex diseases, including male infertility. Firstly, a high-resolution array-CGH platform specific for the X chromosome was used to obtain an outline of X-linked CNVs in infertile patients compared to normozoospermic controls, and led us to the discovery that men with spermatogenic impairment display a higher burden of CNVs in terms of both mean number and mean extension (size in kb). This pilot study also allowed identifying 8 CNVs with potential clinical interest. On one hand, three recurrent deletions, mapping to the Xq27.3 (CNV64) and Xq28 (CNV67 and CNV69), drew our attention for the exclusive (CNV67) or prevalent (CNV64 and CNV69) occurrence in patients. In our follow-up study, six-hundred and twenty-seven idiopathic patients and 628 normozoospermic controls coming from two Mediterranean populations (Spanish and Italian) were further screened for these deletions. The molecular characterization of the deletions' breakpoint was also performed and intersecting functional elements were explored. Our data showed that CNV64 and CNV69 were significantly more frequent in patients than controls ($p=0.013$; $OR=1.89$; 95% CI 1.1 to 3.3, and $p=0.023$; $OR=2.204$; 95% CI 1.05 to 4.62, respectively). CNV69 displayed at least two deletion patterns (type A and type B), of which type B (the larger one) was significantly more represented in patients than controls, thus may account for the potential deleterious effect of CNV69 on sperm production. No genes have been identified within CNV64 and CNV69, nevertheless a number of regulatory elements, have been found to be potentially affected. CNV67 deletion was exclusively found in patients at a

frequency of 1.1% ($p < 0.01$) and a resemblance to the AZFc region on the Y chromosome might be speculated. This deletion may involve the CTA gene *MAGE9A* and/or of its regulatory elements. It may also affect regulatory elements of *HSFX1/2* showing testis-specific expression. Pedigree analyses of two CNV67 carriers indicated that CNV67 deletion is maternally inherited. One of these families was especially informative since, of two brothers, the carrier displayed a pathological semen phenotype whereas his non-carrier brother had a normozoospermic phenotype, strongly suggesting the pathogenic effect of the CNV67 on spermatogenesis. Our second follow-up study focused on the copy number state of the five selected patient-specific gains on the X chromosome and quantitative-PCR on a total of 276 idiopathic infertile patients and 327 controls in a conventional case-control setting and for one interesting locus (intersecting DUP1A) additional 338 subjects were analyzed. This study highlighted a statistically significant difference in the duplication load between patients and controls ($p = 1.65 \times 10^{-4}$). Two of the CNVs were private variants, whereas 3 were found recurrently in patients and never among controls. These CNVs include, or are in close proximity to, genes with testis-specific expression. DUP1A, mapping to the PAR1, was found at the highest frequency (1.4%) that was significantly different compared to controls ($p = 0.047$ after Bonferroni correction). DUP1a includes a long non-coding RNA (*LINC00685*) that potentially acts as a negative regulator of a gene with predicted role in spermatogenesis, *PPP2R3B*. Therefore, we hypothesized that DUP1A may cause spermatogenic failure by affecting the correct regulation of the *PPP2R3B* gene due to a gene dosage-misbalance effect. As an alternative hypothesis, due to its size and location it may disturb X-Y recombination between PAR1 regions during meiosis.

Secondly, we performed a study on the *SHOX* gene in order to evaluate the reliability of the previously reported association between Y-chromosome microdeletions and *SHOX* CNVs, in particular deletions. In our screening, none of the men with microdeletions had *SHOX* deletions, contrasting the aforementioned hypothesis that microdeletions carriers are at a higher risk of developing pathologies caused by *SHOX* haploinsufficiency. These data represent an important contribution to the genetic counseling of infertile couples since our results reassure that the only established risk for ART offspring born from men carrying chromosome microdeletions remains spermatogenic impairment and designates the *SHOX* screening as an utterly unnecessary practice.

From the epigenetic standpoint, a high-resolution methylation microarray was employed to obtain a comprehensive outline of the “normal” sperm DNA methylome, by defining the methylation pattern of >450,000 CpGs in normozoospermic semen samples. Our study,

based on the largest number of subjects (n=8) ever considered for such a great quantity of CpGs, provided data able to determine that the sperm DNA methylation profile is highly conserved among normozoospermic subjects and, for the first time, we demonstrated that this pattern in normozoospermic men remains highly uniform regardless the quality of sperm subpopulations. In addition, our analysis provided both confirmatory and novel data concerning the sperm DNA methylome, including its differential features in relation to somatic and cancer cells.

The widespread use of Assisted Reproductive Technology (ART), which allows infertile men to father children, urges the identification of undiscovered causes and a better understanding of known ones, in order to predict the consequences on reproductive/general health of patients and their future offspring. The work presented in this thesis represents an important progress in the field of idiopathic male infertility. The identification of novel factors that could make a man infertile is of particular importance especially nowadays in the era of ART. Herein, we present data about novel genetic factors associated to spermatogenic impairment that, if confirmed in separate study populations, could lead to a novel genetic test, which may become part of the routine diagnostic work up of infertile men. We also provide novel insights into the sperm DNA methylome providing solid basis for future both basic and clinically oriented research.

2. INTRODUCTION

2.1 Male Infertility

2.1.1. Definition and Etiology

Infertility is defined as the inability of a couple to become pregnant after one year of unprotected intercourse (WHO, 2010). Statistics prove that the incidence of this condition is of about 15% in Western countries: in any given year, about one out of seven couples willing to conceive is not able to do so. Since a male factor is found in half of involuntarily childless couples, it must be assumed that approximately 7% of all men are confronted with fertility problems. Mostly, male factor infertility manifests as abnormal semen parameters –reduced sperm number, motility and morphology- or as alterations of the physical-chemical characteristics of the seminal fluid. The World Health Organization (WHO) established the reference ranges for normal values of sperm parameters, which are included in the recent up-dated version of the guidelines for the examination and processing of human semen (WHO 2010). The possible sperm number alterations are represented by: i) total absence of spermatozoa in the ejaculate where spermatozoa cannot be detected even after centrifugation of the semen sample (azoospermia); ii) reduction of sperm concentration below 15 million spermatozoa/ml (oligozoospermia). Three forms of oligozoospermia can be distinguished: moderate, when sperm concentration is between 15-10 million spermatozoa/ml; severe, when sperm concentration is below 5 million spermatozoa/ml; cryptozoospermia, when spermatozoa are detectable only after cytocentrifugation (in these cases a concentration of <0.01 million spermatozoa/ml is conventionally indicated).

When it comes to motility, the reduction <32% of progressively motile spermatozoa in the ejaculate defines a disorder called asthenozoospermia. As for morphology, we define the condition of teratozoospermia in those cases where the ejaculate presents less than 4% of morphologically normal sperm forms. The majority of infertile patients display anomalies in all three sperm parameters simultaneously and suffer a condition conventionally defined as oligo-astheno- teratozoospermia.

Overall, the etiology of male infertility can be related to a wide range of congenital and acquired factors acting at a pre-testicular, testicular and post-testicular level (Krausz, 2011).

Pre-testicular causes accounts for 10% of infertile forms and are mainly represented by two types of pathological conditions: hypogonadotrophic hypogonadism and coital disorders (erectile dysfunction and ejaculatory disorders, such as *ejaculatio precox* and retrograde ejaculation). Primary testicular dysfunction is the most common cause of spermatogenic impairment (75% of cases) and is related to a number of acquired and congenital etiological factors (testicular causes). A large number of pathologies may lead to an acquired primary testicular failure. Among them are orchitis, testis trauma, torsions, iatrogenic forms (gonadotoxic medications, chemo/radiotherapy, previous inguinal surgery) and some systemic diseases. Anorchia, cryptorchidism (especially bilateral forms) and genetic abnormalities such as karyotype anomalies and Y chromosome microdeletions are well-defined congenital testicular factors of male infertility.

Post-testicular causes represent 15% of cases and include both congenital forms of obstruction/sub-obstruction of the seminal tract- such as congenital absence of the vas deferens (CBAVD)- and acquired forms, which develop from infections/inflammatory diseases of accessory glands or to immunological causes.

The anamnesis, physical examination and semen analysis are of fundamental importance to achieve an accurate diagnosis, thus to reliably orientate the patient to further analyses, therapies or, should these not be applicable, to artificial reproduction techniques (ART).

2.1.2 Genetics of Male Infertility

Diagnostic Features

With the introduction and worldwide diffusion of assisted reproductive techniques (ART) for the treatment of spermatogenesis defects in the male partner, many infertile or subfertile men can now father their own biological children. While males with impaired sperm production due to a genetic factor may now benefit from the wide availability and utilization of ART, the potential risk of transmitting genetic defects to the offspring deserves thoughtful consideration. It is therefore of great importance to detect any genetic anomaly before proceeding to the application of ART.

During the last 30 years, the field of molecular genetics experienced an undeniably pronounced progress with the delivery of new diagnostic tools that allowed the identification of genetic anomalies responsible for spermatogenic impairment. However, known genetic factors collectively explain only 10-15% of all cases of male infertility and despite the advances achieved in the diagnostic workup of the infertile male, the etiopathogenesis of testicular failure remains undefined in about 50% of cases, which are

referred to as “idiopathic”. This lack of understanding significantly limits the ability to counsel patients regarding prognosis for treatment or to optimize personalized treatment strategies for individual patients. In addition, without the identification of the cause for a man’s infertility, it is impossible to tell patients the likelihood of infertility in the offspring that might result from ART. To date, only a limited set of genes is screened in the evaluation of the infertile male and include:

- i) The analysis of the *AR* gene in male with suspected mild form of androgen insensitivity;
- ii) Mutation analysis of a growing number of candidate genes in case of congenital hypogonadotropic hypogonadism, including *KAL1* (Kallmann syndrome 1), *FGFR1* (Fibroblast Growth Factor Receptor 1) o *KAL2*, *FGF8* (Fibroblast Growth Factor 8), *PROK2* (Prokineticin-2), *PROKR2* (G-protein-coupled receptor 54), *GnRH* (Gonadotrophin-releasing hormone), *GnRH-R* (Gonadotrophin-releasing hormone receptor), *GPR54* (G-protein-coupled receptor 54), *LH*, *FSH* and others (Bonomi, 2012).
- iii) The screening for *CFTR* gene mutations in men with congenital absence of the vas deferens (CAVD) without kidney malformations.

The above-mentioned genetic analyses are performed only in selected cases when clear evidence of the associated phenotype exists. When it comes to the routine diagnostic workup of oligo/azoospermic men, only two genetic tests are currently performed: the karyotype analysis for the identification of chromosomal anomalies and the Y-chromosome microdeletion screening (described in paragraph 2.3.2).

Chromosomal Anomalies

Chromosomal aberrations are the consequence of meiotic errors and can be classified as either numerical or structural. Since the first description of an extra X chromosome in association with azoospermia (Ferguson-Smith et al., 1957) many investigations have been performed to determine the contribution of chromosomal anomalies to male infertility. All these studies converge on the hypothesis of a direct relationship between the frequency of chromosomal anomalies and the severity of the testicular phenotype. For instance, patients with <10 million spermatozoa/ml in the ejaculate show 10-fold increased incidence (4%) of carrying autosomal structural abnormalities compared to the general population. Among severe oligozoospermic men (<5 million spermatozoa/ml), the frequency is doubled to 8%, whereas men with non-obstructive azoospermia apparently reach the highest values (15–16%) and abnormalities are mainly related to the sex chromosomes. Klinefelter syndrome

(47, XXY) represents the most common karyotype anomaly in azoospermia and severe male factor infertility. About 80% of patients bear a 47,XXY karyotype whereas the other 20% represented either by 47,XXY/46,XY mosaics or higher grade sex chromosomal aneuploidy or structurally abnormal X chromosome (Krausz, 2011).

A frequent chromosome abnormality associated with azoospermia is attributed to terminal deletions of the long arm of the Y chromosome (Yq-), also visible at the karyotype analysis. These large deletions can derive from the formation of complex structural abnormalities of the Y chromosome, such as the isodicentric (idicYp) and the isochromosome (isoYp) Y chromosome. The idicYp is characterized by the duplication of the short arm (Yp) and of the most proximal region of the Yq- centromere included- and shows the deletion of the terminal part of the Yq. The isoYp is a monocentric Y chromosome (only one centromere is present) showing two Yp and lacking entirely the Yq content. IdicYp and isoYp chromosomes are among the more common genetic causes of severe spermatogenic failure in otherwise healthy men. IdicYp or isoYp formation likely interferes with sperm production via several distinct mechanisms. First, many idicYp and all isoYp chromosomes lack distal Yq genes that play critical roles in spermatogenesis (Skaletsky et al., 2003). Further, idicYp or isoYp formation leads to the duplication of the Yp pseudoautosomal region and deletion of the Yq pseudoautosomal region, which results in the disruption of X-Y meiotic pairing and potentially precludes progression through meiosis (Mohandas et al., 1992). The presence of two centromeric regions makes idicYp chromosomes mitotically unstable. As observed in many human dicentric chromosomes, the mitotic stability of idicYp, especially those with greater intercentromeric distances, is likely to rely upon the functional inactivation of one of the two centromeric regions. However, these chromosomes tend to be lost during mitosis leading to the generation of 45,X cell lines (45, X mosaicism).

As for oligozoospermia, the most frequently found abnormalities are Robertsonian translocations, reciprocal translocations, paracentric inversions and marker chromosomes. The importance of detecting these structural chromosomal anomalies is related to the increased risk of aneuploidy or unbalanced chromosomal complements in the fetus. For example, in the case of Robertsonian translocations, there is a consistent risk of uniparental disomies, which generate through a mechanism called “trisomy rescue” (repairing the trisomic status) during the first division of the zygote.

This correlation between both numerical and structural chromosomal anomalies and impaired sperm production might be related to alterations in the process of chromosome synapsis during meiosis. Studies in mice provided evidence that asynapsed regions may

induce the meiotic checkpoint machinery to eliminate spermatocytes (Odorisio et al., 1998). A similar mechanism might explain why some chromosomal abnormalities in humans are associated with deficient spermatogenesis. Another explanatory mechanism might be related to a structural effect of chromosomal disruptions that might intersect genes involved in spermatogenesis, some of which may be dosage-sensitive and potential mutational targets for chromosomal breakpoints.

State of the Art in Research

Spermatogenesis is a complex process regulated by the concerted action of thousands of genes (Hochstenbach and Hackstein, 2000), but only a small proportion has been hitherto identified and even fewer have been ever analyzed. Therefore, there is a high probability that a wide majority of idiopathic spermatogenic failure has a genetic origin (Krausz and Forti, 2000). The discovery of new genetic associations with male infertility has been hampered by two main factors. First, most studies are underpowered because of insufficient sample size and ethnic/phenotypic heterogeneity. Second, most studies evaluated single candidate genes, a very inefficient approach in the context of male infertility, considering that many hundreds of thousands genes are involved in the process of testicular development and spermatogenesis. The so-called candidate gene approach apparently identified several polymorphisms showing convincing associations with male infertility, but the risk conferred by these polymorphisms is modest and most reported associations either exist as a single study, or they fail to be replicated in subsequent studies; therefore, they did not found any application in the clinical practice (Matzuk and Lamb, 2002; Nuti and Krausz, 2008). However, in spite of the limitations, several robust associations have been identified across multiple studies representing many hundreds of individuals including SNPs in *MTHFR*, *GSTM1* and *FSHB* (Tüttelmann et al., 2012; Wei et al., 2012; Song et al., 2013). The scarcity of informative data derived from studies based on the candidate gene approach in the search of mutations/polymorphisms raised questions about the appropriateness of the used classic screening approach (Nuti and Krausz, 2008). Evidently, the challenge of characterizing the genetic basis of male infertility is certainly a function of the complexity of the spermatogenic process and underlies the significant advantage of whole-genome approaches for discovery research.

The analysis of sequence variants on a genome-wide scale in exceptionally large study populations- defined as Genome Wide Association Study (GWAS) approach- has been used for the identification of genetic factors in several other complex diseases. In the field of male infertility, four genome wide SNP association studies have been performed so far (Aston and

Carrell, 2009; Hu et al., 2012; Kosova et al., 2012; Zhao et al., 2012). The first pilot study, based on a small number of non-obstructive azoospermic (NOA) men and normozoospermic controls of European descent (Aston and Carrell, 2009) and the extended follow up study on 172 SNPs performed by the same group (Aston et al., 2010) provided evidence for some SNPs as potential risk factors and new candidate genes for impaired sperm production. However, these findings have not been confirmed by two subsequent GWAS studies based on exceptionally larger series of NOA subjects (about 1000 cases) and controls (more than 1500 subjects) from the Han Chinese population (Hu et al., 2012; Zhao et al., 2012). Surprisingly, the results from these two studies, based on the same population, do not display any overlap because different genomic regions have been reported as significantly associated with NOA phenotype (1p13.3, 1p36.32 and 12p12.1, by Zhao et al; 6p22 by Hu et al). Finally, in the most recent male infertility GWAS microarray genotyping was performed in 269 men deriving from a founder population of European descent (the Hutterites) that desires large families and proscribes contraception (Kosova et al., 2012). Forty-two SNPs associated with reduced fertility traits in the Hutterites were evaluated in 123 ethnically different infertile men and only nine SNPs were significantly associated with sperm quality and/or function (Kosova et al., 2012). The divergence between the GWAS studies mentioned above might be only partially related to ethnic differences, since overlapping candidate SNPs could not be found even in the two large GWAS based on the same Chinese population. The inability of GWAS approach in identifying relevant SNPs involved in spermatogenic failure may actually depend on the fact that SNP arrays are based on common genetic variants. For instance, it has been predicted that it is more likely that “rare” variants rather than common polymorphisms are involved in the etiology of spermatogenic impairment (Aston and Carrell, 2009). Another explanation for the inconsistency and unsuccessful outcome of GWA studies may be that the pathogenic effect of SNPs is related to the combination of low size effect SNPs or their interaction with the environment. In this regard, a significant association with male infertility has been clearly demonstrated for the 677C>T polymorphism in the *MTHFR* (5,10-Methylene-tetra-hydrofolate reductase) gene in populations with low folate intake (Nuti and Krausz, 2008).

A more recent strategy to identify genetic causes of male infertility has been the employment of genomic microarrays to identify Copy Number Variations (CNVs). The study of CNVs was initially applied to a number of multifactorial complex diseases and allowed the identification of novel genetic factors involved in the etiology of some types of cancer, neurological and autoimmune diseases or susceptibility to HIV-1 infections. Being infertility

indeed a complex disease, it is plausible that CNVs affecting regions or multi-copy genes relevant to spermatogenesis may also contribute to infertility. To date, the only CNVs that have been demonstrated to be in a clear-cut cause-effect relationship with male infertility are the AZF microdeletions on the Y chromosome.

2.1.3. Epigenetics Aspects

Male infertility is a complex condition where not only genes, but also epigenetic factors are predicted to play an important role. Epigenetics is defined as *the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in the DNA sequence*. These “epigenetic” changes encompass an array of molecular modifications of DNA or histones, which are intimately associated with the DNA. Epigenetic factors not only have a profound impact on developmental processes but also have relevance in many diverse areas of biology and medicine, including cancer biology, the study of environmental effects, and the study of aging (Herceg and Vaissière, 2011; Liu and Rando, 2011). For instance, epigenetic mechanisms regulate chromatin accessibility throughout an organism’s lifetime, as specific sets of genes are active at any stage of development. Each cell type has its own epigenetic signature that reflects the developmental history and environmental influences, and is ultimately reflected in the phenotype of the cell and organism.

The sperm epigenetic program is unique. First, a major erasure of epigenetic marks occurs in primordial germ cells (PGCs). Second, male germ cell nuclei undergo a reorganization and a condensation of their genome during post-meiotic maturation, including DNA methylation, histone acetylation, implementation of histone variants, and histone-to-protamine transition during spermiogenesis. Epigenetic signals are crucial for the proper functioning of the sperm genome, and any alteration during these processes may contribute to altering sperm function and fertilization efficiency.

Human spermatozoa undergo three types of epigenetic modifications:

- Extensive chromatin modifications as a result of the removal of histones during spermiogenesis and their replacement with protamines;
- Chemical modifications seen in histones retained in sperm chromatin;
- The methylation to the sperm DNA itself.

All these are potent modulators of gene transcription, including the transcriptional regulation during embryogenesis. However, for the purposes of this thesis only sperm DNA methylation will be discussed in detail.

In this thesis, the term *epigenome* will be used to indicate the whole set of epigenetic modifications that can occur in the genome and the term *methylome* will refer specifically to methylation changes.

Sperm DNA methylation

DNA methylation is a potent epigenetic mark that mainly promotes gene silencing, is essential to allele-specific imprinting of certain genes, and is crucial in X chromosome inactivation (Ng et al., 1999; Bronner et al., 2007). DNA methylation occurs by covalent attachment of a methyl group to the C5 position of cytosine bases found in cytosine-phosphate-guanine dinucleotides (CpGs) of eukaryotic DNA. This conversion to 5-methylcytosine is possible due to action of DNA methyltransferase (DNMT) enzymes. Different members of the DNMT family of enzymes act either as *de novo* DNMTs, putting the initial pattern of methyl groups in place on a DNA sequence, or as maintenance DNMTs, copying the methylation from an existing DNA strand to its new partner after replication (Portela and Esteller, 2010). Mammalian genomes are punctuated by DNA sequences containing an atypically high frequency of CpG sites termed CpG islands (CGIs). These conspicuous unique sequences are approximately 1 kb in length and overlap the promoter regions of 60–70% of all human genes (Weber et al., 2007). CGIs have been shown to co-localize with the promoters of all constitutively expressed genes and approximately 40% of those displaying a tissue restricted expression profile (Larsen et al., 1992; Zhu and Yao, 2009). Consistent with this promoter association, CGIs are generally characterized by a lack of DNA methylation, thus by a transcriptionally permissive chromatin state. However, there is a percentage, though small, of CGIs that acquires methylation during normal development, and some of these examples are known to play a key role in X-chromosome inactivation and genomic imprinting (Handy et al., 2011; Liu et al., 2011).

As anticipated, studies on mice allowed defining that mammalian development is accompanied by two major waves of genome-wide demethylation and re-methylation: one during germ-cell development and the other after fertilization (Fig. 2.1). Upon fertilization, the zygotic genome experiences a broad reprogramming process, which contributes to the transition of the zygote into a totipotent state and during which epigenetic reprogramming crucially regulates zygotic gene expression. For instance, paternal DNA is actively demethylated at the one-cell stage, followed by passive loss of methylation that reaches a minimum at the blastocyst stage. Conversely, maternal DNA is maintained in a hypermethylated state at the one-cell stage and is passively demethylated in a replication-dependent manner. Hence, at the blastocyst stage, embryos possess a globally low level of

DNA methylation. However, upon implantation, the epiblast and early primordial germ cells (PGCs) restore a hypermethylated state due the action of *de novo* DNMTs. Subsequently, after migration and proliferation of PGCs, a second wave of DNA demethylation is detected. Following gender determination, *de novo* DNA methylation takes place and new methylation landscapes are established in an asymmetrical pattern in male and female germ cell precursors. In male germ cells, *de novo* methylation initiates before the onset of meiosis, and finishes prior to birth. In female germ cells, instead, *de novo* methylation occurs during the postnatal development of meiotic prophase I - arrested oocyte (Fig. 2.1).

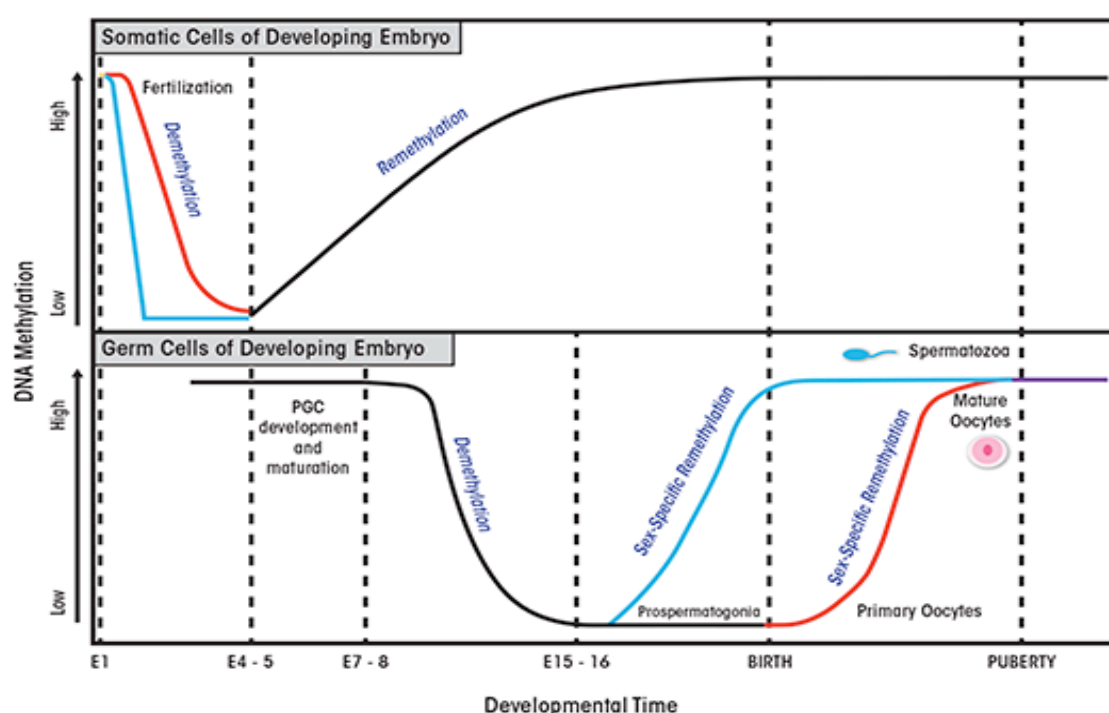


Figure 2.1. Reprogramming in mammalian development. Two waves of epigenetic reprogramming occur during embryo development. The first phase of reprogramming occurs in the normal body cells (i.e., somatic cells) of the developing embryo. In mice, following fertilization, the embryo undergoes genome-wide demethylation that is completed by embryonic day 5 (E5). The paternal genome (blue line) undergoes rapid, active demethylation, whereas in the maternal genome (pink line), demethylation occurs via a passive process. Remethylation of the embryonic genome begins at day E5 and is completed prior to birth. The second wave of epigenetic reprogramming occurs in the germ cells of the developing embryo, which will ultimately give rise to gametes that contain sex-specific epigenetic signatures. The primordial germ cells (PGCs) of the developing embryo contain the methylation signatures of the parental genomes. At approximately E7–8, the PGCs undergo rapid demethylation that is complete by E15–16. Following this, sex-specific methylation is re-established. In the male germline, reprogramming is complete at birth (blue line), whereas in females, reprogramming continues until puberty (violet line). Figure from (Ungerer et al., 2013).

To date, still little is known about the role of sperm methylation in the epigenetic regulation of embryonic development, but this issue is progressively becoming clarified. For instance, studies exploring the sperm methylome in relation to embryogenesis have been published

already (Hammoud et al., 2009; Molaro et al., 2011) and agree that developmental gene promoters are generally in a hypomethylated state, supporting the hypothesis of a potential embryonic role for the sperm genome. Although these studies provide indirect evidence only, they sustain a potential epigenetic hypothesis of epigenetic programming in spermatozoa, in which hypomethylation of regulatory regions - together with simultaneous activating and silencing histone modifications - poises, but not yet actively marks, key developmental genes for rapid activation in development, while the rest of the genome is either silenced by protamine binding or bound to histones with silencing modifications.

Sperm DNA methylation and Male Fertility

Early evidence for a link between epigenetic markers and male fertility was demonstrated in mice by the administration of a demethylating agent, the 5-aza-20-deoxycytidine. Reduced sperm production was observed in association with reduction of testis and epididymal weight, reduced littermate size, and higher neonatal mortality (Kelly et al.; Doerksen, 1996). A number of studies have explored the relationship between DNA methylation levels and male fertility in humans. The first association between methylation defects and infertility was reported by Benchaib et al. (Benchaib et al., 2005), who reported that global methylation levels above an arbitrary threshold was seemingly linked to high pregnancy rates. Another study reported for the first time that poor-quality semen samples displayed abnormally elevated levels of DNA methylation of both imprinted and non-imprinted genes and other repetitive elements (Houshdaran et al., 2007). The authors proposed that these elevated methylation levels arose from of an improper erasure of methylation marks in cases of oligo-astheno-teratozoospermia (OAT) rather than from *de novo* methylation following epigenetic reprogramming. The analysis of 23,094 CpG probes performed by (Aston et al., 2012) allowed the identification of three individuals - two men with abnormal protamine 1/protamine 2 ratio and one patient resulting in abnormal embryogenesis after IVF/intracytoplasmic sperm injection treatment - displaying broad disruption of sperm DNA methylation profiles, thus suggesting that disruptions in sperm DNA methylation may be an important mark in some infertile men. Considering the importance of genomic imprinting acquisition during spermatogenesis, Marques et al. compared the methylation patterns of *H19* (paternally methylated) and *MEST* (paternally unmethylated) imprinted genes in spermatozoa of fertile and oligozoospermic infertile men, by bisulphite genomic sequencing (Marques et al., 2004, 2008). Oligozoospermic patients displayed an erroneous loss of methylation of the *H19* gene and such decrease was associated with a decrease in sperm count. Furthermore, an unexpected *MEST* hypermethylation was observed in these patients

(Marques et al., 2008). These findings were confirmed by a Japanese study investigating the DNA methylation status of seven imprinted genes: *H19* and *GTL2* (paternally methylated), and *PEG1*, *LIT1*, *ZAC*, *PEG3* and *SNRPN* (maternally methylated) in infertile patients. Also they found that oligozoospermic cases presented decreased methylation levels of the paternal DNA methylation in *H19* and *GTL2* and increased methylation levels in *PEG1*, *LIT1*, *ZAC*, *PEG3* and *SNRPN*. Subsequently, Boissonnas et al. (2010) quantitatively compared the methylation levels at 47 different CpGs included in *IGF2* and *H19* genes between normozoospermic and oligozoospermic men (Boissonnas et al., 2010). This study showed that these two genes were abnormally hypomethylated in oligozoospermic men and that this drastic loss of methylation correlated with the severity of the oligozoospermia. The association between OAT and methylation errors, both hypermethylation and hypomethylation, were subsequently confirmed by other research groups even when different techniques were used (Hammoud et al., 2010; Poplinski et al., 2010; Sato et al., 2011; Aston et al., 2012).

Finally, some groups studied the methylation state of the promoter of some genes involved in the spermatogenic process. Patterns of abnormal methylation associated with OAT were reported in the *DAZL* promoter (Navarro-Costa et al., 2010c) and the *CREM* promoter (Nanassy and Carrell, 2011). Moreover, aberrant hypermethylation of the *MTHFR* promoter was found in 53% of patients with non-obstructive azoospermia (Khazamipour et al., 2009) and in some cases of idiopathic infertility (Wu et al., 2010). All these studies support the hypothesis that sperm DNA methylation patterns of both imprinted and non-imprinted genes are essential for normal sperm function, fertility, and embryo development. However, the etiology and whether methylation errors are acquired during fetal or early post-natal development are still unresolved questions.

Given that epigenetic signals are crucial for the proper functioning of the genome, phenotypic differences in sperm production (quantitative and qualitative traits) both at an inter- and intra-individual level, existing even within a context of normal spermatogenesis, may reflect a variation in the sperm epigenome. The only study available in this regard evaluated intra- and inter-individual DNA methylation changes in normozoospermic men by the analysis of 12,198-CpG sites (Flanagan et al., 2006). The authors reported significant intra- and inter- individual variations for 6 genes, supporting the hypothesis that there is a link between epigenetic variations and the variability of semen phenotypes. Contrastingly, two later studies reported only a limited inter-individual variability in DNA methylation in normozoospermic men, though it must be considered that comparison was performed only

between two sperm donors and using different techniques -methylated DNA immunoprecipitation (MeDIP) procedure and promoter arrays (Hammoud et al., 2009) and a genome-wide shotgun bisulfite sequencing (Molaro et al., 2011).

Another issue scarcely investigated so far is whether different semen fractions containing qualitatively different spermatozoa display different epigenetic landscapes. The only data available in this regard are provided in a study by Navarro-Costa et al (2010), in which significant differences in the methylation pattern of the *DAZL* promoter were observed between normal and defective germ cell fractions deriving from the same individual.

An attempt of filling this gap of knowledge in relation to methylation patterns in quality-fractionated semen samples was performed in our laboratory, and the study performed will be presented in this thesis.

2.2. Copy Number Variations (CNVs)

2.2.1. *Discovery and Definition*

The discovery of Copy Number Variations (CNVs) dates back to 2002 when Charles Lee, in his attempt of genotyping a group of patients, started finding that healthy controls showed major variations in their genetic sequences, with some having more copies of specific genes than others. Lee began to collaborate with Steven Scherer, who had made similar observations, and together through array-based comparative genomic hybridization (a-CGH) approaches measured for the first time the occurrence of these copy variants across the genome. Meanwhile, Michael Wigler was also observing differences in copy numbers in healthy individuals using a complementary microarray technique involving representational oligonucleotide probes to detect amplifications and deletions in the genome. Finally, in 2004, Lee's and Scherer's groups of researchers published the first evidences that large-scale variations in copy number were common and occurred in hundreds of loci in the human genome, including areas coding for disease-related genes (Iafrate et al., 2004; Sebat et al., 2004). By then, a CNV was defined as a DNA segment of one kilobase (kb) or larger that is present at a variable copy number in comparison with a reference genome. The term "variation" instead of "polymorphism" is used because the relative frequency of most CNVs in the general population have not yet been well defined and the term polymorphism is limited to genetic variants that have a minor allele frequency of $\geq 1\%$ in a given population.

2.2.2. What are CNVs?

CNVs are a class of structural variation, which includes also balanced alterations regarding position and orientation of genomic segments defined as translocations and inversions, respectively. The term CNV is not generally used to indicate variations caused by insertion/deletion of transposable elements.

These unbalanced quantitative variants can be classified into:

- **Gains** when an increase of genetic material is observed compared to the reference genome as a consequence of duplication/amplification or insertion events. The amplified DNA fragments can be found adjacent to (tandem duplication) or distant from each other and even on different chromosomes.
- **Losses** when a reduction or the complete loss of genetic material is observed compared to the reference genome as a consequence of deletion events. In the present thesis the terms “loss” and “deletion” will be used to indicate the reduction and the complete loss (null genotype) of a given DNA sequence compared to the reference genome, respectively.

A CNV can be simple in structure or may involve complex gains or losses of homologous sequences at multiple sites in the genome (Fig.2.2).

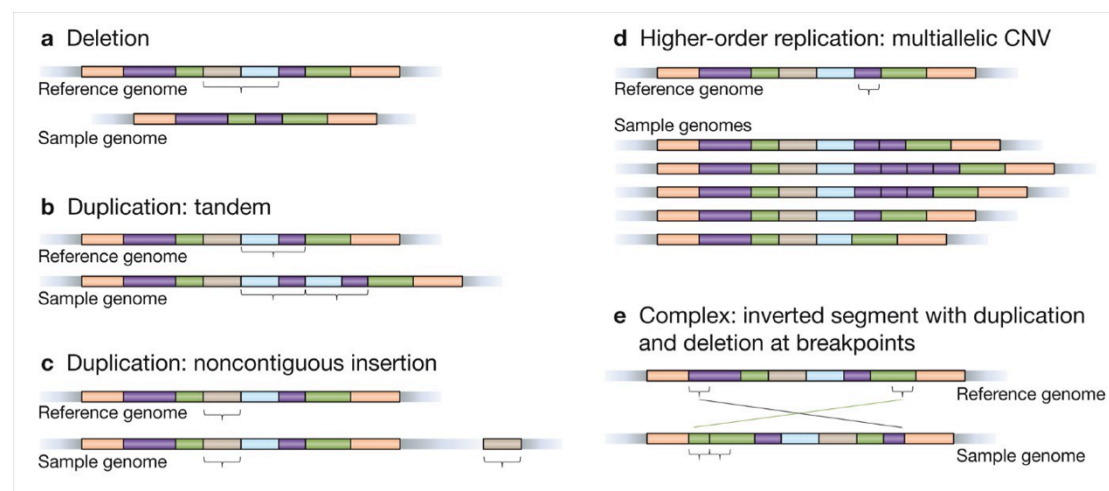


Figure 2.2. Different types of CNV. CNVs (in the sample genome) are defined by comparison with a reference genome. DNA blocks displaying sequence identity are represented with the same color. a) Deletion of two contiguous fragments (deletion); b) Tandem duplication (gain); c) Duplication (gain) with insertion of the duplicated sequence far from the origin; d) Multiallelic gain produced by multiple duplication event; e) Complex CNVs resulting from inversion, duplication and deletion events. Figure from (Lee and Scherer, 2010).

In 2006, a collaboration of international research laboratories built the first genome-wide map of existing CNVs in the human genome (Redon et al., 2006). An examination of 270 DNA samples from the multiethnic population employed by the HapMap Project revealed a total of 1447 discrete CNVs. Taken together, these CNVs cover approximately 360 Mb, i.e. 12% of the human genome, with a prevalence of small rearrangements (<20 Kb). The HapMap Project notes that CNVs encompass more nucleotide content per genome than SNPs, underscoring CNVs' significance to genetic diversity. The map of CNVs shows that no region of the genome is exempt, and that the percentage of an individual's chromosomes that exhibit CNVs varies anywhere from 6% to 19% (Redon *et al.*, 2006). The genomic regions encompassed by these CNVs contain hundreds of genes and functional elements and many CNVs reached a population frequency greater than 1% (Copy Number Polymorphisms). These observations, together with the inter-individual variability in gene copy number (Redon et al., 2006; Jakobsson et al., 2008), lead to hypothesize the importance of CNVs in the evolutionary process and in the adaptation to diverse environmental conditions. Indeed, CNVs are an important genetic component of phenotypic diversity (Wong et al., 2007), and represent the primary source of inter-individual variability between genomes (Iafrate et al., 2004; Sebat et al., 2004; Redon et al., 2006).

With the growth of information on CNVs in the human genome, the accurate annotation of these structural variations has become progressively important. At present, several databases are currently available for genome-wide investigation of genomic variants, the most important of which is the "Database of Genomic Variants" (<http://dgv.tcag.ca/dgv/app>), which provides a comprehensive continuously up-dated catalog of the structural variations identified the human genome. For each CNV several information are annotated: whether it is a gain or a loss, the exact genomic position, the frequency and bibliographic references to trace back to the study that produced those data. To date (December 2014) a total of 353126 CNVs deposited by a total of 62 studies.

2.2.2. Potential Mechanisms of CNV Generation

The rate of CNV formation is estimated to be several orders of magnitude higher than any other type of mutation and the molecular mechanisms by which they generate seems to be similar in bacteria, yeast and humans. *De novo* formation of CNVs can occur in both the germline and somatic cells. Bruder et al. (Bruder et al., 2008) provided evidence for the possible generation of CNVs during mitosis (in somatic cells) by reporting that monozygotic twins show different CNVs at different loci. These CNVs presumably arose during early stages of embryogenesis, immediately before or immediately after embryonic division in the

two individuals. It is therefore plausible that some CNVs might originate during embryogenesis even in the case of a single pregnancy, generating a “chimerism” for such CNVs within the same individual; this phenomenon has been also demonstrated by (Piotrowski et al., 2008), who observed the presence of CNVs, affecting a single organ or one or more tissues of the same subject. Other evidences for the onset of CNVs at the somatic level are the presence of CNV mosaicism in tumor tissues (Fridlyand et al., 2006; Darai-Ramqvist et al., 2008) as well as in blood cells of healthy subjects (Lam and Jeffreys, 2006, 2007).

CNVs often occur in regions reported to contain, or be flanked by, large homologous repeats or segmental duplications (SDs)(Fredman et al., 2004; lafrate et al., 2004; Sharp et al., 2005; Tuzun et al., 2005). SDs (also referred by some as low copy repeats or LCRs; (Lupski, 1998) can be defined as duplicated DNA fragments that are >1 kb and map either to the same chromosome or to different, non-homologous chromosomes (Bailey et al., 2002; Lupski and Stankiewicz, 2005). Segmental duplications could arise by tandem repetition of a DNA segment followed by subsequent rearrangements that place the duplicated copies at different chromosomal loci. Alternatively, segmental duplications could arise via a duplicative transposition-like process: copying a genomic fragment while transposing it from one location to another (Eichler, 2001).

According to whether CNVs are associated with segmental duplications or not, they may be susceptible to structural chromosomal rearrangements via two general mechanism, respectively:

- Homologous recombination-based pathways including the non-allelic homologous recombination mechanism (NAHR);
- Non-homologous recombination-based pathways, including non-homologous end-joining (NHEJ) and the Fork Stalling and Template Switching (FoSTeS) models.

Certain CNVs may be found to be associated with non-beta DNA structures (DNA regions that differ in structure from the canonical right-handed alfa-helical duplex, including left-handed Z-DNA and cruciforms). Such DNA structures are believed to promote chromosomal rearrangements (Kurahashi and Emanuel, 2001; Bacolla et al., 2004) and may also theoretically contribute to the genesis and maintenance of certain CNVs. There may be a relationship between the size of a given CNV and its associated mutational mechanism(s). For example, data from at least two studies have shown that larger CNVs are more frequently associated with segmental duplications than smaller CNVs (Fig.2.3), which are predominantly generated by non-homology-driven mutational mechanisms.

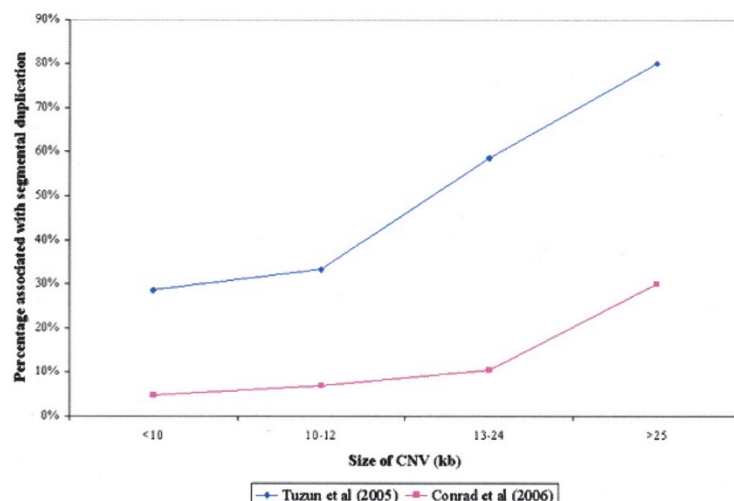


Figure 2.3. Graph showing the positive correlation between the size of CNVs and the likelihood of association with SDs. This correlation is noted by both the Conrad et al. (2006) and Tuzun et al. (2005) studies.

NON-ALLELIC HOMOLOGOUS RECOMBINATION MECHANISM

NAHR is driven by the extended sequence homology between two region of the genome oriented in the same direction- such as the above mentioned SDs (Shaw and Lupski, 2004; Stankiewicz and Lupski, 2010)- where incorrect pairing during meiosis/mitosis or DNA repair across homologous regions can result in a gain or loss of intervening sequence. Homologous recombination underlies several mechanisms of accurate DNA repair, where another identical sequence is used to repair a damaged sequence. If a damaged sequence is repaired using homologous sequence in the same chromosomal position within the sister chromatid or in the homologous chromosome (allelic homologous recombination) no structural variation will occur. However, if a crossover forms when the interacting homologies are in non-allelic positions on the same chromosome or even on different chromosomes this will result in an unequal crossing-over causing the duplication and subsequent deletion of the intervening sequence. More specifically, inter-chromosomal and inter-chromatid NAHR between LCRs with the same orientation results in reciprocal duplication and deletion, whereas intra-chromatid NAHR creates only deletions (Fig. 2.4).

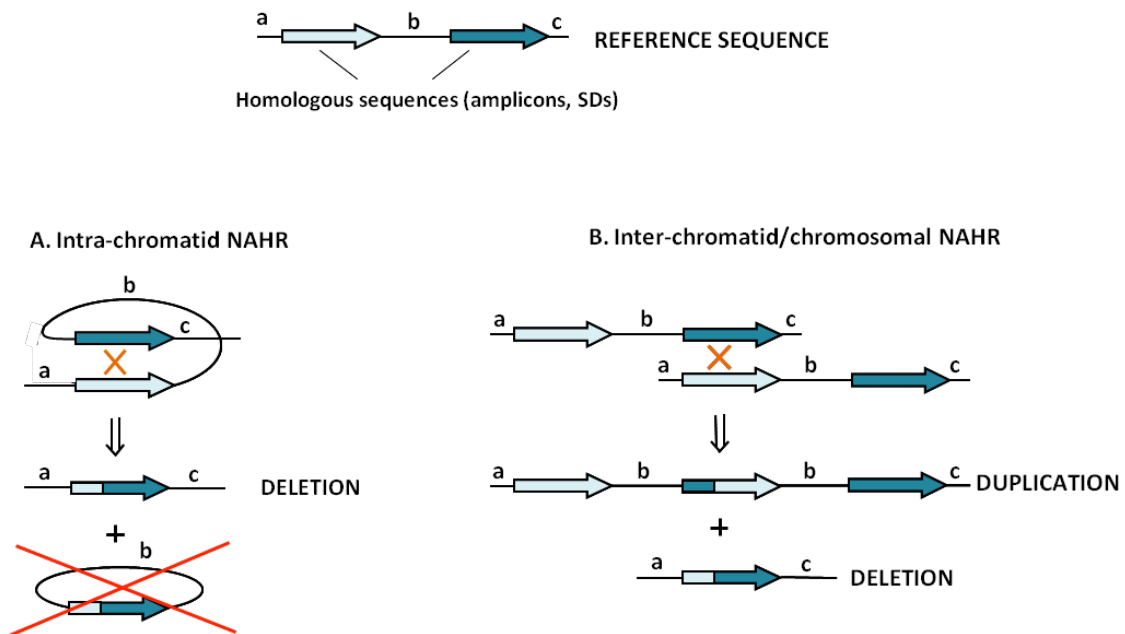


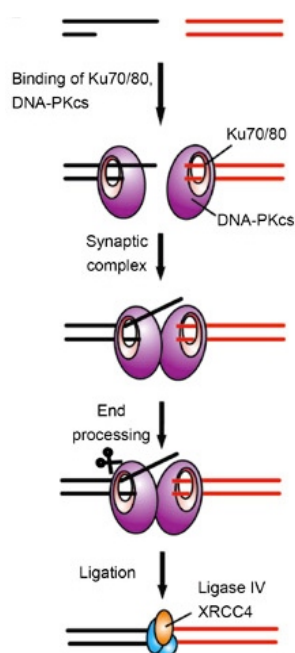
Figure 2.4. NAHR mechanisms. Recombination occurs between two directly oriented SDs represented by yellow and blue arrows. Two scenarios are possible: A. Intra-chromatid NAHR: recombination between two homologous sequences on the same chromatid results in the deletion of the interposed DNA segment. B. Inter-chromatid or inter-chromosome NAHR: two non-allelic homologous sequence on sister chromatids or chromosomes are involved in recombination leading to a deletion and the reciprocal duplication.

Theoretically, the frequency of deletions should be always higher than that of duplications. However, if deleterious deletions underwent negative selection, duplications would then occur at a higher frequency (Turner et al., 2008). Therefore, duplication frequency should not exceed deletion frequency, unless negative selection in both germ cells and somatic cells makes deleterious deletions very rare or not represented.

NON HOMOLOGOUS END-JOINING (NHEJ)

NHEJ is one of the two major mechanisms used by eukaryotic cells to repair DNA double strand breaks (DSBs) without involving a template DNA sequence. This non-homologous DNA repair pathway has been described in organisms from bacteria to mammals [66-68 in Gu et al] and is routinely used by human cells to repair both 'physiological' DSBs, such as in V(D)J recombination, and 'pathological' DSBs, such as those caused by ionizing radiation or reactive oxygen species. NHEJ proceeds in four steps (Fig. 2.5): detection of DSB; molecular bridging of both broken DNA ends; modification of the ends to make them compatible and ligatable; and the final ligation step (Weterings and van Gent, 2004). Being a non-homology based mechanism, NHEJ does not require DNA pairing for successful ligation and, consequently, unlike NAHR does not depend on the presence of SDs. Evidence exists that

NHEJ is more prevalent in unstable (or fragile) regions of the genome such as the sub-telomeric regions (Nguyen et al., 2006; Kim et al., 2008). Furthermore, many NHEJ events,



classified as microhomology-mediated end joining, require end resection and join the ends by base pairing at microhomology sequences (5–25 nucleotides)(McVey and Lee, 2008; Pawelczak and Turchi, 2008). NHEJ leaves a “molecular scar” since the product of repair often contains additional nucleotides at the DNA end junction (Lieber, 2008).

Figure 2.5. NHEJ brings the ends of the broken DNA molecule together by the formation of a synaptic complex, consisting of two DNA ends, two Ku70/80 and two DNA-PK_{CS} molecules. Non-compatible DNA ends are processed to form ligatable termini, followed by repair of the break by the ligase IV/XRCC4 complex. Figure from (Weterings and Chen, 2008).

Fork stalling and template switching (FoSTeS)

Study of stress-induced amplification of the lac genes, using the *E. coli* Lac system by Cairns and (Cairns and Foster, 1991), led (Slack et al., 2006) to propose that template switching was not confined to a single replication fork, but could also occur between different replication forks. This model, now called fork stalling and template switching (FoSTeS), illustrated in Figure 2.6, proposes that when replication forks stall in cells under stress, the 3' primer end of a DNA strand can change templates to single-stranded DNA templates in other nearby replication forks. This hypothesis was necessary because the mean length of amplicons in that study was about 20kb (Slack et al. 2006), which is too long to have occurred within a replication fork. According to this model, during DNA replication, the DNA replication fork stalls at one position, the lagging strand disengages from the original template, transfers and then anneals, by virtue of microhomology at the 3' end, to another replication fork in physical proximity (not necessarily adjacent in primary sequence), 'primes', and restarts the DNA synthesis (Lee et al., 2007). The invasion and annealing depends on the microhomology between the invaded site and the original site. Upon annealing, the transferred strand primes its own template driven extension at the transferred fork. This priming results in a 'join point' rather than a breakpoint, signified by a transition from one segment of the genome to another – the template-driven juxtaposition of genomic sequences. Switching to

another fork located downstream (forward invasion) would result in a deletion, whereas switching to a fork located upstream (backward invasion) results in a duplication.

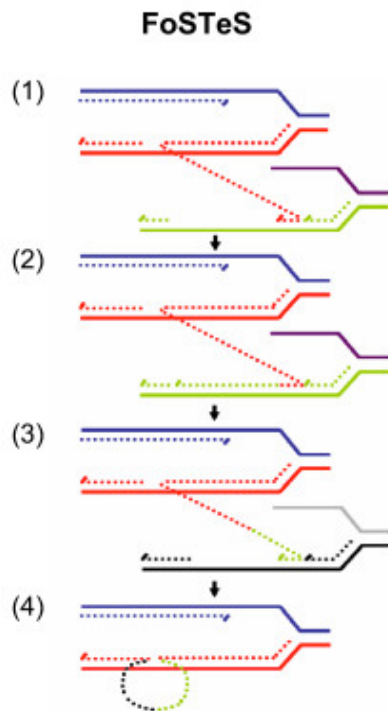


Figure 2.6. After the original stalling of the replication fork (dark blue and red, solid lines), the lagging strand (red, dotted line) disengages and anneals to a second fork (purple and green, solid lines) via microhomology (1), followed by (2) extension of the now 'primed' second fork and DNA synthesis (green, dotted line). After the fork disengages (3), the tethered original fork (dark blue and red, solid lines) with its lagging strand (red and green, dotted lines) could invade a third fork (gray and black, solid lines). Dotted lines represent newly synthesized DNA. Serial replication fork disengaging and lagging strand invasion could occur several times (e.g. FoSTeS x 2, FoSTeS x 3, etc.) before (4) resumption of replication on the original template. Figure from (Gu et al., 2008).

2.2.3. CNVs and Disease: Functional Consequences

As already mentioned, CNVs are a common feature of the genome of healthy humans and rather play an important role in evolution and adaptation to different environments, as major source of genetic inter-individual variability (Iafrate et al., 2004; Sebat et al., 2004). However, the gain or loss of DNA sequence can also produce a spectrum of functional effects and human disease phenotypes. One obvious way by which CNVs might exert their effect is by altering transcriptional levels (and presumably subsequent translational levels) of the genes that are in variable copy number. CNVs may influence gene function directly by altering the copy number of dosage-sensitive genes or by disrupting the gene-coding sequence: partial gain or loss of coding sequences can produce different alleles, including both loss and gain of function. For example, deleted internal exons could result in a frameshift and subsequent loss of function through truncation or non-sense mediated decay. Chimeric proteins can also be produced when CNV breakpoints lie within two different genes, leading to the fusion of two partial coding regions. CNVs in non-coding regions can also lead to several position effects via the deletion or transposition of critical regulatory elements, such as promoters, enhancers and silencers or disrupting the function of these, leading to

changes in sequence or location with respect to a target gene (Hurles et al., 2008). Apparently, the functional effect of a CNV is strictly dependent on the exact position of the CNV breakpoint, i.e. the region where a fragment was inserted (gain) or lost (loss/deletion). The main consequences through which CNVs may act are represented in Figure 2.7. CNV size is scarcely predictive of the phenotypic effect, since a number of apparently benign CNVs are of an order of magnitude of 2 Mb, and in some cases can also reach a 10 Mb length (Redon et al., 2006; Hansson et al., 2007). Although the functional consequences of a CNV might be difficult to predict, many CNVs do generate alleles with a clear-cut impact on health. For instance, the development of new high-resolution tools- such as genome-scanning array technologies and comparative DNA-sequence analyses- CNVs have been associated with a growing number of common complex diseases, including human immunodeficiency virus (HIV), autoimmune diseases such as Chron disease, psoriasis, systemic lupus erythematosus (Aitman et al., 2006; Fanciulli et al., 2007; Willcocks et al., 2008; Bassaganyas et al., 2013), a spectrum of neuropsychiatric disorders as autism, schizophrenia (Cook and Scherer, 2008; Rodríguez-Santiago et al., 2010; Saus et al., 2010) and some type of cancer (neuroblastoma, breast and prostate cancer).

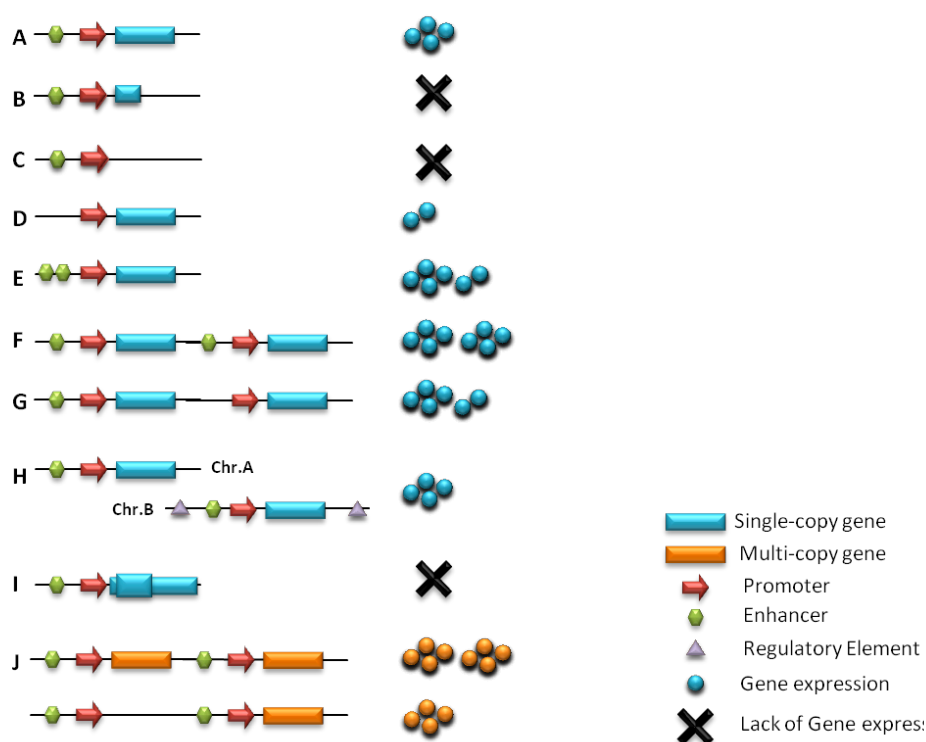


Figure2.7. Impact of CNVs on gene expression. A. Single copy dosage-sensitive gene (reference genome): promoter, upstream enhancer element and coding sequence are represented; partial and complete deletion affecting coding sequence (B and C); Deletion and duplication affecting enhancer (D and E); Complete and partial (not involving enhancer) tandem duplication (F and G); Complete tandem inter-chromosome duplication involving a regulatory element inhibiting gene expression (H); Partial tandem duplication disrupting coding sequence (I); multicopy gene loss (J).

Being infertility indeed a complex disease, researchers involved in this field started investigating whether CNVs might play a role in the etiopathogenesis of spermatogenic failure: an increased number or a specific distribution of CNVs could result in defective recombination, meiotic failure and loss of germ cells or CNVs might affect the activity of individual genes important for spermatogenesis. To date, 6 studies have been published on the relationship between CNVs and male infertility (Tüttelmann et al., 2011; Krausz et al., 2012a; Stouffs and Lissens, 2012; Lopes et al., 2013)(Tüttelmann et al., 2011; Krausz et al., 2012a; Stouffs et al., 2012; Lopes et al., 2013), and all converge on the hypothesis that infertile patients have a significantly higher burden of CNVs in their genome compared to normozoospermic controls. This issue was the main objective of this thesis and will, thus, be presented subsequently.

2.3 The Y Chromosome

2.3.1. Structure and gene content

The Y chromosome is a submetacentric chromosome and with its 60 Mb of length is one of the smallest chromosomes of the human genome. The Y chromosome is peculiar in its structure, which can be conceptually divided in two genomic regions: i) the male-specific region of the Y (MSY); ii) the pseudoautosomal regions (PARs), which correspond to the domain of X–Y homology involved in meiotic pairing (Fig.2.8).

The MSY region, comprising approximately 95% of the chromosome length, lacks a homologous region on the other sex chromosome, thus it is genetically isolated from meiotic recombination. Within the MSY region both euchromatic and heterochromatic DNA sequences can be identified: the euchromatic portion covers approximately 23 Mb of the chromosome, including 8 Mb on the short arm (Yp) and 14.5 Mb on the long arm (Yq); apart from the large block of centromeric heterochromatin of approximately 1 Mb, (Tyler-Smith et al., 1993) typically found in every nuclear chromosome, the Y chromosome also harbors another, much longer heterochromatic block (40 Mb) that encompasses the distal part of the Yq.

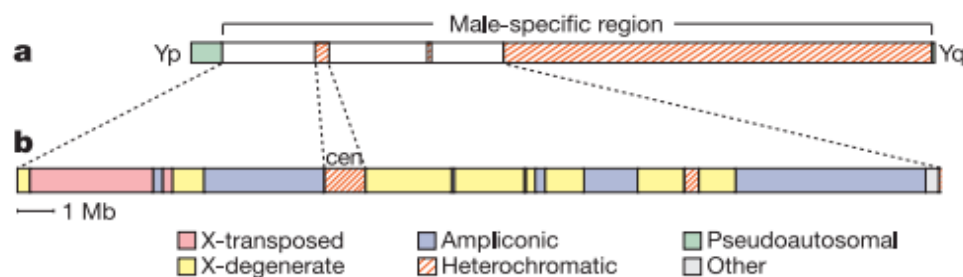


Figure 2.8. Schematic representation of the whole Y chromosome, including the pseudoautosomal MSY regions (from Skaletsky 2003). Heterochromatic segments and the three classes of euchromatic sequences (X-transposed, X-degenerate and Amplificonic) are shown.

MSY's euchromatic portion can be divided into three classes, firstly defined by (Skaletsky et al., 2003):

- **X-transposed sequences:** are the result of a massive X-to-Y transposition occurred about 3–4 million years ago, after the divergence of the human and chimpanzee lineages (Page, 1987; Mumm et al., 1997; Schwartz et al., 1998). Subsequently, an inversion within the MSY short arm cleaved the X-transposed block into two non-

contiguous segments, as observed in the modern MSY. Found only on the short arm of the Y chromosome, these sequences show 99% homology to some sequences in Xq21; notwithstanding they do not participate in X–Y crossing over during male meiosis. Within the X-transposed segments, which have a combined length of 3.4 Mb, only two genes can be identified. In fact, the X-transposed sequences display the lowest density of genes among the three classes as well as the highest density of interspersed repeat elements (in particular LINE).

- **X-degenerate sequences:** in contrast to the X-transposed sequence blocks, these segments are dotted with single-copy gene or pseudogene homologues of 27 different X-linked genes. These single-copy MSY genes and pseudogenes display between 60% and 96% nucleotide sequence identity to their X-linked homologues, and they seem to derive from ancient autosomes from which the X and Y chromosomes co-evolved. Among these is also present the *SRY* (Sex Determining Region) gene, which has an important role in sex determination and is the only X-degenerate sequence to be expressed predominantly in the testes.

Amplicons: are large DNA blocks that exhibit marked similarity—as much as 99.9% identity over tens or hundreds of kilobases—to other sequences in the MSY. These sequences are located in seven segments scattered across the euchromatic long arm and proximal short arm of the Y chromosome with a combined extension of 10.2 Mb. Amplicons, which can be regarded to as SDs, are in turn organized in symmetrical arrays of contiguous units named “palindromes” (Fig.2.9 below).

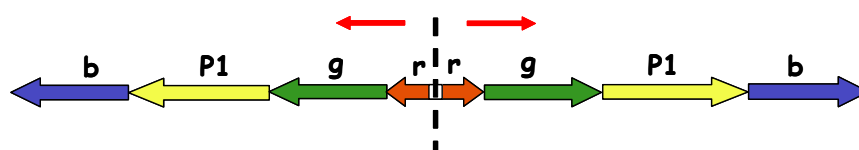


Figure 2.9. Example of organization of the amplicons (colored arrows) in a symmetrical array of continuous repeat units (palindrome P1)

Eight massive palindromes can be identified in the ampliconic region of the Yq, each defined by a symmetry axis separating two largely identical arms (with a sequence identity of 99.94–99.997%) constituted by single or multiple amplicons (Kuroda-Kawaguchi et al., 2001; Skaletsky et al., 2003) (Fig. 2.9). Amplicons represent approximately 35% of the MSY region and the eight palindromes collectively cover one quarter of the MSY euchromatin. Consequentially, the Y chromosome displays a significantly higher SDs content compared to the rest of chromosomes showing an average content of approximately 5%. The ampliconic

sequences exhibit by far the highest density of genes, both coding and non-coding, among the three sequence classes in the MSY euchromatin. Nine multi-copy distinct MSY-specific protein-coding gene families can be identified, with copy numbers ranging from two (*VCY*, *XKRY*, *HSFY*, *PRY*) to three (*BPY2*) to four (*CDY*, *DAZ*) to six (*RBMV*) to on average 35 (*TSPY*). Overall, these nine coding multi-copy gene families encompass approximately 60 transcription units and are predominantly or exclusively expressed in the testis. Furthermore, the ampliconic sequences include at least 75 other transcription units for which strong evidence of protein coding is lacking. Of these 75 putative non-coding transcription units, 65 are members of 15 MSY-specific families, and the remaining 10 occur in single copy. Such an abundance of multiple copies and palindromes have led researchers to call the Y chromosome a "hall of mirrors." Although this sequence repetition created great challenges in the sequencing of the Y chromosome, the complex structure also serves an important purpose. Multiple copies of essential spermatogenesis genes ensure that in spite of deletion events, which may result in the loss of a single copy of an essential gene, spermatogenesis can still proceed via proteins produced by remaining copies.

The ampliconic sequences evolved from a great variety of genomic sources, and accumulated in the MSY region through two main molecular mechanisms: the amplification of X-degenerate genes and the transposition/retroposition and subsequent amplification of autosomal genes. Such an accumulation of SDs (amplicons) in the MSY region is considered an evolutionarily conserved strategy of the Y chromosome to counteract the accumulation of deleterious mutations, in the absence of conventional recombination with a chromosome partner. The presence of massive near-identical amplicons allows indeed two recombination-based DNA repair mechanisms to occur in the MSY region: gene conversion and non-allelic homologous recombination (NAHR). The first is a non-reciprocal transfer of sequence information from one DNA duplex to another (Szostak et al., 1983), which can occur between duplicated sequences on a single chromosome and in mitosis (Jackson and Fink, 1981). Gene conversion (non-reciprocal recombination) in the MSY is as frequent as crossing over (reciprocal recombination) is in ordinary chromosomes, and occurs routinely in 30% of the MSY (Skaletsky et al. 2003). This conversion-based system of gene copy "correction" permits the preservation of Y-linked genes from the gradual accumulation of deleterious mutations ensuring their continuity over time. As stated above, NAHR is a homology-based mechanism of accurate DNA repair, which can also lead to the generation of large-scale AZF structural rearrangements such as inversions and CNVs affecting the dosage of a number of Y-linked genes.

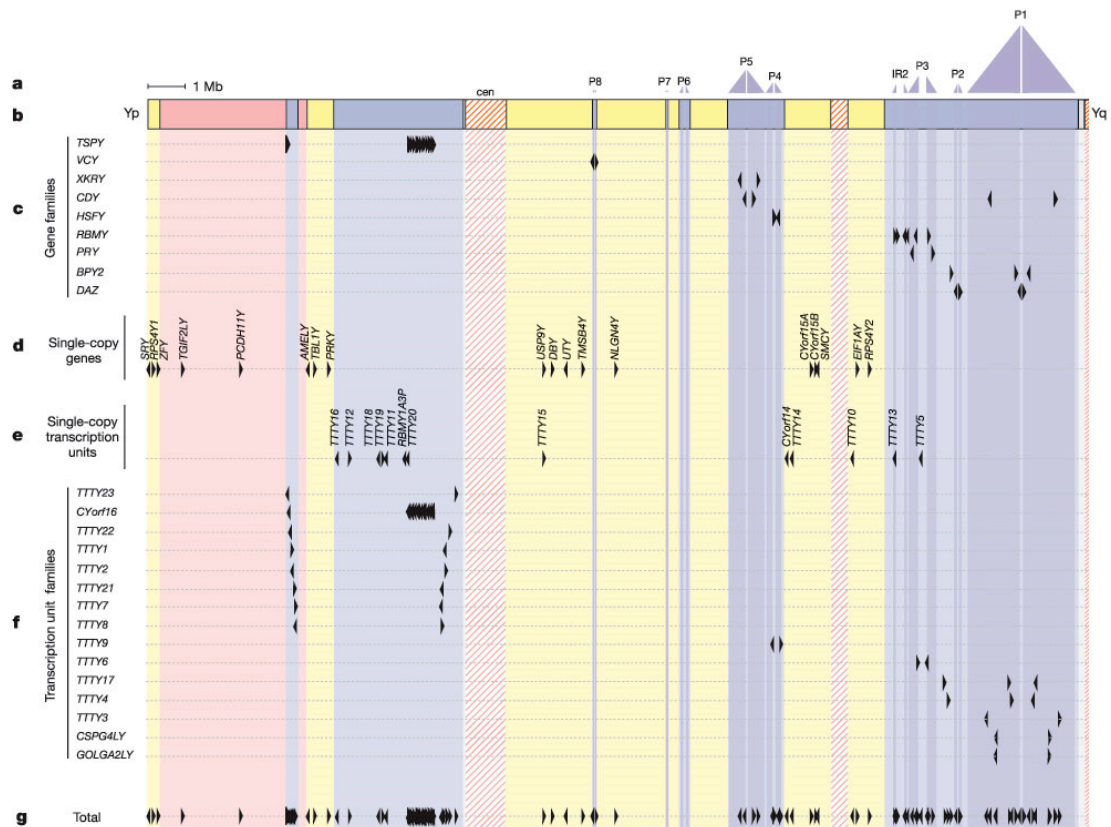


Figure 2.9. MSY genes, transcription units and palindromes. a) Localization of the 8 palindromes (P1-P8) in the MSY region. b) MSY region as represented in Figure 2.8. Solid black triangles denote coding genes and transcription units which are classified as follow: c) Nine families of protein-coding genes; d) single-copy protein coding genes and e) single-copy transcription units which give rise to spliced but apparently non-coding transcripts. f) Fifteen families of transcription units. g) Merged map of all genes and transcription units. Figure from Skaletsky et al. 2003.

The MSY is flanked on both sides by **pseudoautosomal regions (PAR1 and PAR2)** short regions of homology between the mammalian X and Y chromosomes. Being present on both sex chromosomes, the PARs act like autosomes and recombine during meiosis. Thus, genes in this region characterized by autosomal inheritance rather than a strictly sex-linked fashion. PAR1 comprises 2.6 Mb of the short-arm tips of both X and Y chromosomes and is required for pairing of the X and Y chromosomes during male meiosis. All characterized genes within PAR1 escape X inactivation. X-Y pairing in the PAR is thought to serve a critical function in spermatogenesis, at least in humans and mouse (Burgoyne et al., 1983; Matsuda et al., 1992; Mohandas et al., 1992). PAR2 is located at the tips of the long arms and is much shorter, spanning only 320 kb (Freije, 1992). PAR2 exhibits a much lower frequency of pairing and recombination than PAR1 and is not necessary for fertility (Helena Mangs and Morris, 2007).

Thanks to the first sequencing of the human X chromosome (Ross et al., 2005), it is known that PAR1 contains at least 24 genes, of which 16 have been hitherto characterized (Table 2.1), whereas 4 genes (*SPRY3*, *SYBL1*, *IL9R* and *CXYorf1*) have been identified on the PAR2.

Table 2.1. PAR1 Genes and Protein Function. Only characterized genes are showed.

Gene	Description	Function
<i>PLCXD1</i>	Phosphatidylinositol-Specific Phospholipase C, X Domain Containing 1	Ubiquitously expressed; diverse biological functions including roles in inflammation, cell growth, signaling and death and maintenance of membrane phospholipids.
<i>GTPBP6</i>	GTP Binding Protein 6 (Putative)	Ubiquitously expressed. Undefined function.
<i>PPP2R3B</i>	Protein Phosphatase 2, Regulatory Subunit B	Exerts regulatory control over the initiation of DNA replication. Overexpression causes G1 cell cycle arrest.
<i>SHOX</i>	Short Stature Homeobox	Homeobox-containing gene, thought to be a transcription factor related to short stature syndromes.
<i>CRLF2</i>	Cytokine receptor-like factor 2	Receptor for TSLP, a cytokine that enhances the maturation process of dendritic cells and promotes the proliferation of CD4+ T cells.
<i>CSF2RA</i>	Colony-stimulating factor 2 receptor, alpha	Alpha subunit of the receptor for the granulocyte-macrophage colony-stimulating factor (GM-CSF). GM-CSF is important for the growth and differentiation of eosinophils and macrophages in the bone marrow, and also regulates cell viability in human embryos.
<i>IL3RA</i>	Interleukin 3 receptor, alpha	Alpha subunit of the receptors for interleukin 3.
<i>SLC25A6</i>	Solute carrier family 25, member A6	Member of the ADP/ATP translocase family, which has a potential role in Th cell survival and immune cell homeostasis.
<i>ASMTL</i>	Acetylserotonin O- methyltransferase-like	Undefined function.
<i>P2RY8</i>	Purinergic receptor P2Y, G-protein coupled, 8	Member of the purine nucleotide G-protein coupled receptor gene family
<i>AKAP17A</i>	A Kinase (PRKA) Anchor Protein 17A	Splice factor regulating alternative splice site selection for certain mRNA precursors. Mediates regulation of pre-mRNA splicing in a PKA-dependent manner
<i>ASMT</i>	Acetylserotonin O-methyltransferase	Catalyzes the final reaction in the synthesis of melatonin.
<i>DHRSXY</i>	Dehydrogenase/reductase (SDR family) X-linked	Encodes an oxidoreductase of the short-chain dehydrogenase/reductase family.
<i>ZBED1</i>	Zinc finger, BED-type containing 1	Binds to 5'-TGTCG[CT]GA[CT]A-3' DNA elements found in the promoter regions of a number of genes related to cell proliferation. Binds to the histone H1 promoter and stimulates transcription.
<i>CD99</i>	CD99 molecule	Cell surface molecule involved in T-cell adhesion processes. Activation of a distinct domain of CD99 activates a caspase-independent death pathway in T-cells.
<i>XG</i>	Xg blood group	Encodes the XG blood group antigen.

Genes in **bold** were object of investigation of this thesis.

Disturbances of PAR1 have been linked to clinical conditions, such as isolated short stature, Leri–Weill dyschondrosteosis and Langer mesomelic dysplasia, all associated with mutations in the *SHOX* gene.

PAR1 and Disease: the SHOX gene

The *SHOX* (Short stature HOmeoboX-containing) gene, for the first time isolated by positional cloning (Rao et al.), covers a genomic region of 40 kb and resides in the PAR1 of human sex chromosomes at Xp22 and Yp11.3, within a 170-kb region, 500 kb from the telomeres. Because genes residing in the PAR1 escape X-inactivation, two copies of the *SHOX* gene are normally expressed in males as well as in females. The *SHOX* gene has one non-coding and six coding exons (ranging from 58 to 1166 bp in size) and encodes for two isoforms of a homeodomain-containing transcription factor, SHOXa (293 amino acids) and SHOXb (225 amino acids), which differ in the C-terminal region. These two different isoforms are expressed in differential tissues: SHOXa is expressed at low levels in many tissues, while SHOXb has a more restricted expression, displaying the highest expression in bone marrow fibroblasts (Clement-Jones et al., 2000; Rosenfeld, 2001; Munns et al., 2004). The involvement of the SHOX protein in bone development is further supported by two studies reporting its expression in hypertrophic chondrocytes of the growth plate (Marchini et al., 2004; Munns et al., 2004). Furthermore, Marchini and colleagues (2004) provided evidence that the expression of *SHOX* also induces cell cycle arrest and apoptosis in osteosarcoma cells as well as in primary chondrocytes, suggesting an involvement of the protein in the processes regulating chondrocyte differentiation.

SHOX haploinsufficiency - a condition that results when one copy of a dosage-sensitive gene has been deleted - was suggested for the first time to be involved in idiopathic short stature (ISS; OMIM# 604271) and the short stature in Turner syndrome (TS) by two consequent publications (Ellison et al., 1997; Rao et al., 1997), whereas homozygous loss of the gene has been correlated with the Langer type mesomelic dysplasia (OMIM# 249700). It has also been shown that *SHOX* haploinsufficiency can cause not only short stature but also Turner skeletal abnormalities such as short fourth metacarpals, *cubitus valgus* and characteristics of Léri-Weill dyschondrosteosis (LWD; OMIM #127300) (Kosho et al., 1999). Subsequently, heterozygous *SHOX* mutations were also shown to cause Léri-Weill dyschondrosteosis (Belin et al., 1998). It is overall estimated that the incidence of *SHOX* deficiency is between 1/2000-1/5000 in the general population and 1/40-1/150 among short people. In contrast to the clear picture existing concerning the loss of one (haploinsufficiency) or both copies of *SHOX* and its deleterious effects on stature, poor and unclear information is available with respect to *SHOX* over-dosage, reported in association with normal to tall stature (Ogata et al., 2001; Adamson et al., 2002) as well as to more severe conditions such as LWD and idiopathic short stature (ISS) (Roos et al., 2009; Thomas et al., 2009; D'haene et al., 2010; Benito-Sanz et al.,

2012). *SHOX* duplications apparently are quite rare, with only few cases hitherto reported (Grigelioniene et al., 2001; Tachdjian et al., 2008; Roos et al., 2009; Thomas et al., 2009; D’haene et al., 2010; Gervasini et al., 2010) and no direct relationship with any specific phenotype has yet been defined. To date, there is only one study in the literature reporting an association between spermatogenic impairment and *SHOX* CNVs (Jorgez et al., 2011), in which the authors propose that the mechanism underlying Y-chromosome microdeletions (paragraph 2.3.2) might also lead to the formation of PAR rearrangements, including *SHOX* deletions and duplications.

2.3.2. *Y chromosome-linked CNVs*

The Y chromosome is peculiar for its haploid nature that precludes recombination with the X chromosome for most of its length. This unique characteristic led to Y-Y recombination events between homologous sequences within the Y chromosome. This phenomenon led to the accumulation of a relevant number of segmental duplications (also called amplicons), which constitute a favorable substrate for CNV formation because of their susceptibility to both NAHR and gene conversion, previously described in paragraph 2.3.1.

Since that the Y chromosome does not undergo recombination during the crossing-over, a phylogenetic approach was used to study the dynamics of Y-linked CNVs and their mechanisms of formation (Jobling, 2008). This study showed that determining the frequency of a given CNV in different Y lineages allows deducing the minimum number of independent mutation events accounting for the CNV distribution. As illustrated in Fig. 2.10, the dynamics of the Y-linked CNVs can be ascribed to:

- **Unique CNVs:** these are present in all the members of a given Y haplogroup but absent in other lineages (CNV1 in the Fig.2.10). In this case, the mutation is a unique event that has occurred in the ancestral Y chromosome of that specific haplogroup.
- **Recurrent CNVs:** distributed among different branches, may arise through several independent mutation events reflecting the highly mutagenic nature of the involved region (CNV2 and CNV3 in the Fig.2.10). In the case of recurrent CNVs showing high prevalence in Y haplogroups (CNV2), belonging to more than one lineage indicates that the mutation has likely occurred in the ancestral Y chromosome of more than one lineage; though, in some members of the same haplogroup “reversion” of the mutation has occurred. This mainly occurs in cases of CNVs with a high mutation rate. Finally, CNVs that occur with very high recurrence can also form as independent events in different Y lineages (CNV3).

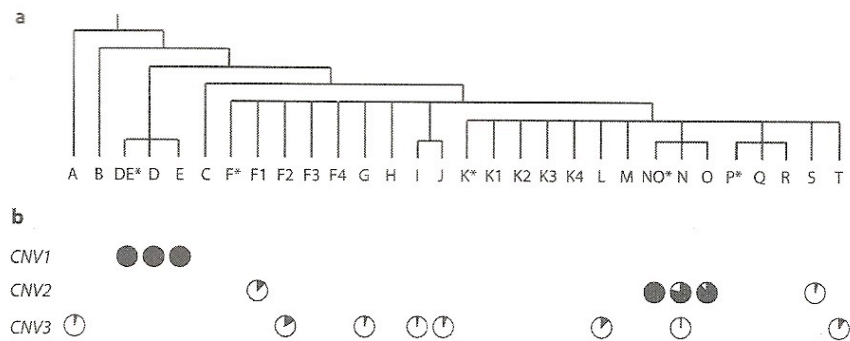


Figure 2.10. Phylogenetic approach used for the study of the dynamics of Y-linked CNV formation. Figure from Joblin et al 2008.

Furthermore, making an estimate of the number of generations encompassed by a sampled chromosome during evolution allows inferring the mutation rate of a certain CNV (Hammer and Zegura, 2002; Repping et al., 2006; Karafet et al., 2008).

Y-linked CNVs have been object of investigation in several fields of research, from forensic and population genetics studies to molecular reproductive genetics, the latter being mainly focused on the investigation the AZF region in men with altered spermatogenesis. The comprehensive and objective picture nowadays available in relation to Y-linked CNVs is the fruit of groundbreaking studies that provided systematic genome-wide CNV surveys (Redon et al., 2006; Perry et al., 2008) and re-sequencing data of the entire whole Y chromosome (Levy et al., 2007). The largest scale study (Redon et al. 2006) employed CGH to explore 104 distinct Y chromosomes from the HapMap sample (n=270), revealing that the AZFc region is the most variable euchromatic portion in terms of CNVs (Fig. 2.11.). In the following paragraph the more common known AZF-linked CNVs will be described.

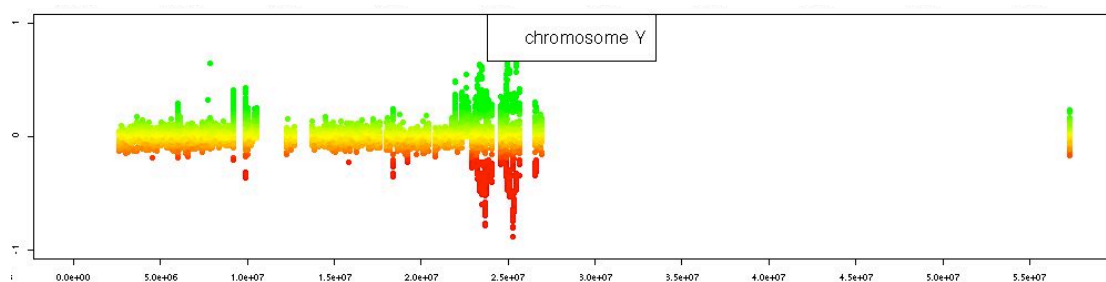


Figure 2.11. Representation of the log2 ratio from comparative genomic hybridization to BAC clones spanning the Y euchromatin. The most dynamic region corresponds to the AZFc region. (Figure by Jobling et al. 2008).

AZF microdeletions

The first association between azoospermia and microscopically detectable deletions in the long arm of the Y chromosome (Yq), was reported by Tiepolo and Zuffardi in 1976 (Tiepolo and Zuffardi, 1976). The authors proposed the existence of an AZoospermia Factor (AZF) on

Yq, representing a key genetic determinant for spermatogenesis, since its deletion was associated with the lack of spermatozoa in the ejaculate. Due to the structural complexity of the Y chromosome, the molecular characterization of the AZF took about 30 years to be achieved. With the development of molecular genetic tools and the identification of specific markers on the Y chromosome (Sequence Tagged Sites, STSs), it was possible to circumscribe the AZF region. Three AZF sub-regions were identified in proximal, middle and distal Yq11 and designated AZFa, AZFb and AZFc, respectively. It was later demonstrated that AZFb and AZFc overlap, being 1.5 Mb of the distal portion of AZFb interval part of the AZFc region (Fig. 2.12).

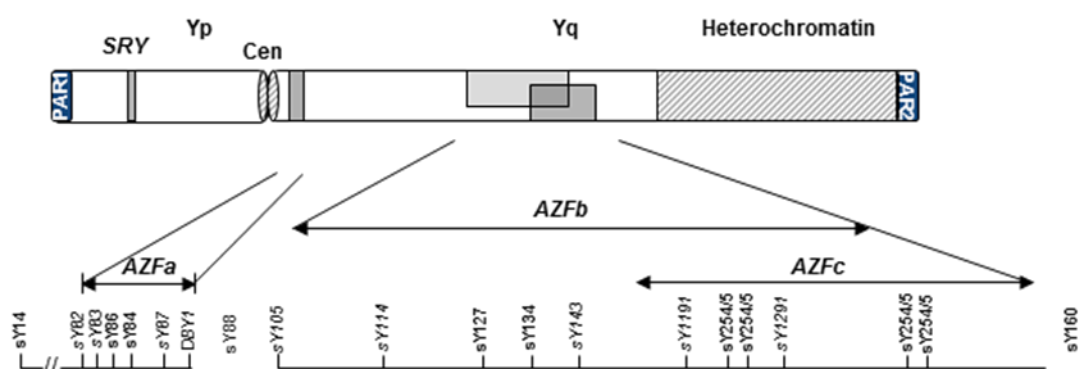


Figure 2.12. Structure of AZF region of the Yq. Schematic representation of the Y chromosome showing the three AZF regions (A) with each specific STSs (B).

Sub-microscopic deletions involving the AZF regions, regarded to as Y-chromosome microdeletions, occur in 1/4000 males in the general population. They are now considered a well-established genetic cause of male infertility being exclusively found in men with impaired sperm production. The AZF microdeletion screening is now part of the diagnostic work-up of severe male factor infertility (Krausz et al., 2014). Indications for AZF deletions screening are based on sperm count and include azoospermia and severe oligozoospermia (<5 million spermatozoa/ml). Thanks to the EAA guidelines and EAA/EMQN external quality control program (<http://www.emqn.org/emqn/>) Yq testing has become more homogeneous and reliable in different routine genetic laboratories. The EAA guidelines provide a set of primers (two markers for each region), which is able to detect virtually all clinically relevant deletions. Deletion frequency increases with the severity of spermatogenic impairment amounting to 5-10% among non-obstructive azoospermic men, 2-4% among severe oligozoospermic men with less than 5 million spermatozoa/ml and less than 0.5% among men with sperm concentration between 5 and 10 million spermatozoa/ml. Y-chromosome microdeletions have never been found among men with normal sperm parameters

(normozoospermic men). In addition to the semen phenotype, also ethnic background is likely to influence the occurrence of this genetic anomaly as suggested by the different deletion frequency observed even within similar semen categories amongst infertile men from different populations. In this regard, the lowest deletion frequency (1.8%) was reported in German and Danish idiopathic severely oligozoospermic men (Cruger et al., 2003; Simoni et al., 2008), whereas the highest in an ethnically admixed population from France (13.7%) (Krausz et al., 1999) and in Romanians (10%) (Raicu et al., 2003). In most of the cases Y microdeletions arise as *de novo* event and are associated with heterogeneous seminal and testicular phenotypes.

Y microdeletions arise through NAHR and, according to their recombination hot-spot, they can be classified as AZFa, P5-proximal P1 (AZFb), P5-distal P1 (AZFbc), P4-distal P1 (AZFbc) and b2/b4 (AZFc) deletions.

The AZFa region spans 792 Kb and unlike either AZFb or AZFc, is exclusively constituted by single-copy DNA. The complete deletion of AZFa interval results from non-allelic homologous recombination between two flanking HERV elements (human endogenous retroviral elements), spanning 10 Kb each and displaying an overall 94% of sequence identity. Two ubiquitously expressed genes map inside the AZFa region and are thus involved in the deletion: *USP9Y* and *DDX3Y*. The AZFa deletion is a rather rare event- less than 5% of the reported Y microdeletions (Kamp et al., 2001; Krausz and Degl'Innocenti, 2006)- and it is invariably associated with azoospermia due to the complete absence of germinal cells in seminiferous tubules, a condition known as pure Sertoli Cells Only Syndrome (SCOS) (Kleiman et al., 2012). The low prevalence most likely depends on both limitations of the deletion mechanism (it is characterized by a relatively short recombination target), and the potential negative selection of the deletion due to its deleterious effect on fertility. The corresponding NAHR product, the AZFa duplication, is detected at a fourfold higher frequency when compared to that of the deletion indicating that increased AZFa gene dosage does not affect fertility (Bosch and Jobling, 2003).

The AZFb region spans a total of 6.23 Mb and contains three single-copy regions, a DYZ19 satellite repeat array and 14 ampliconic elements organized in palindromes (from P2 to P5 and the proximal part of P1) of which P5/P4 and P1 are the NAHR targets giving rise to the complete and partial AZFb deletion, respectively. AZFb deletion carriers are azoospermic with testicular histology of maturation arrest at the spermatocyte/spermatid stage. Unlike the AZFa deletion, no evidence for reciprocal duplications have been reported for the AZFb deletion, so far.

The AZFc deletion spans 3.5 Mb and results from the NAHR between the flanking b2 and b4 amplicons. The deletion removes 21 genes and transcriptional units belonging to 8 multicopy gene families (Fig.19). These include 3 protein coding gene families (*BPY2*, *CDY* and *DAZ*) specifically expressed in the testis. The AZFc deletion, accounting for approximately 60% of all recorded AZF deletions (Navarro-Costa et al., 2010a), is associated with severe spermatogenic impairment phenotype (azoospermia or severe oligozoospermia) related to variable testicular pictures ranging from pure and mixed SCOS to hypospermatogenesis and maturation arrest. A deterioration of semen quality over time has been suggested for AZFc deleted oligozoospermic men based on indirect observations such as the difference in age between carriers with azoospermia and oligozoospermia or the increase of FSH concentrations over time in some subjects. However, this issue is nowadays still debated.

Within the AZFc region three different patterns of partial deletions have been identified, the *gr/gr*, *b2/b3* and *b1/b3* deletions (Repping et al., 2006) (Fig. 2.13) but only the *gr/gr* deletion is of potential clinical interest. Four meta-analyses are published on this topic and all report that the *gr/gr* deletion confers on average a 2- to 2.5-fold increased risk of reduced sperm output/infertility (Tüttelmann et al., 2007; Visser et al., 2009; Navarro-Costa et al., 2010a; Stouffs et al., 2011) making this deletion a unique example in andrology of a confirmed significant genetic risk factor for impaired sperm production. The *gr/gr* deletion removes half of the genetic content (1.6 Mb) of the AZFc region. Eight testis-specific gene and transcription unit families are affected by this deletion pattern. In particular, it removes two copies of the *DAZ* gene and 1 copy of *CDY1* gene, which are the two most important AZFc candidate infertility genes.

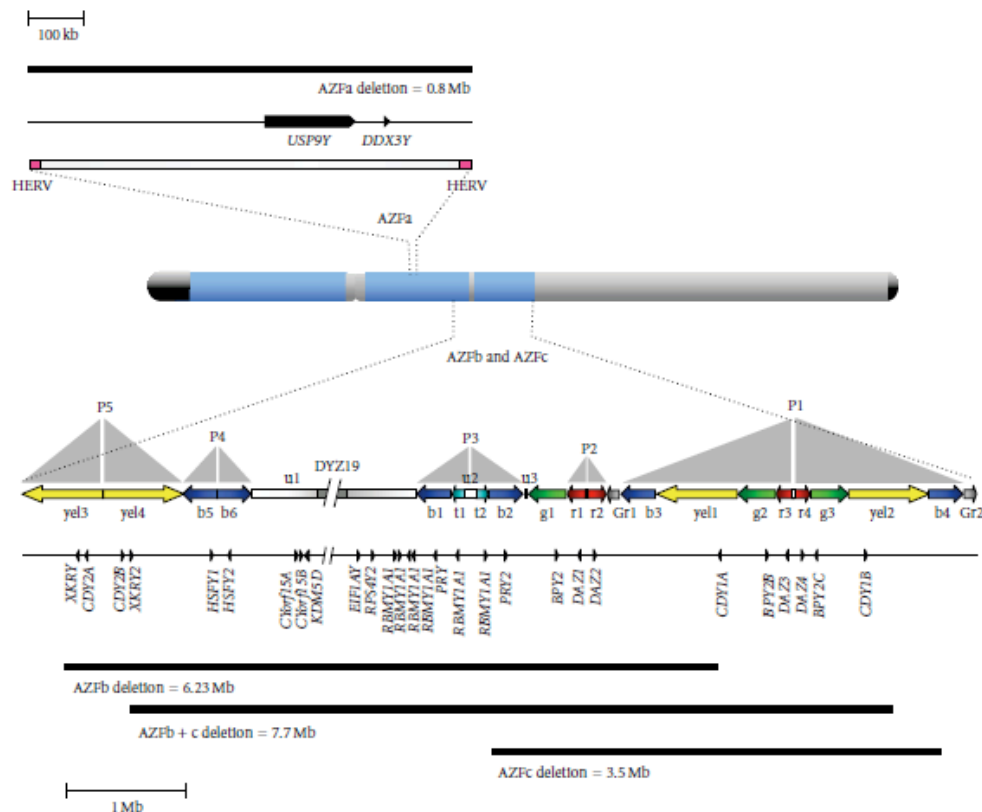


Figure 2.13. AZF deletion patterns. Recombining amplicons /palindromes responsible for each AZF deletions and genes involved are shown. *AZF_a* is flanked by two human endogenous retrovirus (HERV) elements that mediate the occurrence of *AZF_a* deletions via non-allelic homologous recombination. *AZF_b* and *AZF_{b+c}* deletions are caused by P5/proximal P1 yel3/yel1) and P5-distal P1 (yel3/yel2) recombination, respectively. NAHR between b2 and b4 amplicons lead to *AZF_c* deletion. Figure from Navarro –Costa 2010a.

The identification of Yq microdeletions, which explain the etiology of the impaired spermatogenesis, is not only relevant from a diagnostic standpoint, but it also has a prognostic value prior testicular biopsy (TESE) (Brandell et al., 1998; Krausz et al., 2000b). In this regard, in case of complete *AZF_a* and *AZF_b* deletions of the Y chromosome testicular biopsy is not advised because the chance of finding spermatozoa is virtually zero. The *AZF_c* deletion is compatible with the presence of spermatozoa in the testis or in the ejaculate, and is obligatorily transmitted to the male offspring. Therefore, genetic counseling for infertile couples willing to undertake ART treatment is mandatory. The severity of spermatogenic failure in the son may vary considerably, although given the strict cause–effect relationship between AZF deletions and impaired spermatogenesis, normal spermatogenesis cannot be warranted. When it comes to the exact testicular phenotype, predictions cannot be made because of the different genetic background and environmental factors that will have impacted on the reproductive functions and fertility potential of the father and his son.

Concerns have been raised about the potential risk for Turner's syndrome (45,X) in the offspring and other phenotypic anomalies associated with sex chromosome mosaicism, including ambiguous genitalia. Data on men with Y microdeletions (Siffroi et al., 2000; Rajpert-De Meyts et al., 2011) and in patients bearing a mosaic 46,XY/45,X karyotype with sexual ambiguity and/or Turner stigmata (Patsalis et al., 2005) suggest that some Yq microdeletions are associated with an overall Y-chromosomal instability which might result in the formation of 45,X cell lines. A recent study (Jorgez et al., 2011) reported that 5.4%% of men with AZF deletions and a normal karyotype also carried *SHOX* haploinsufficiency. Indeed, this study raised the question about the importance of screening for *SHOX*-linked CNVs in men carrying Y-chromosome microdeletions, which constitutes one of the objects of this thesis.

2.4. The X chromosome

2.4.1. General features and structure

The X chromosome is a sub-metacentric chromosome representing almost 5% of the total DNA in women, whereas in men, who are hemizygous for the X chromosome, it represents about 2.5% of the total DNA. The reference sequence of the human X chromosome has been recently reassembled (Mueller et al., 2013) and accordingly its total size amounts to about 155.3 Mb. It displays a low (G+C) content (39%) compared with the genome average (41%) and it is highly enriched in repetitive sequences. These regions account for 56% of the euchromatic X-chromosome sequence and are represented by:

- **Short Interspersed Nuclear Elements (SINEs)** belonging to the *Alu* family, the content of which in the X chromosome is below the genome average.
- **Long Terminal Repeats (LTRs)** the coverage of which is above average.
- **Long Interspersed Nuclear Elements (LINEs)** of the L1 family, which are the most represented class of repetitive elements of the X chromosome, accounting for 29% of the chromosome sequence compared to a genome average of only 17% (Ross et al., 2005).
- **Ampliconic sequences** (segmental duplications of >10 Kb sharing > 99% nucleotide identity) represent approximately 2% (3.15 Mb) of the chromosome length (Mueller et al., 2013).

The cross-species alignment of orthologous X-linked genes allowed defining two evolutionary domains that are characteristic of the X chromosome:

- The **X-conserved region (XCR)**, an ancestral region including all the long arm and PAR1, which would descend from the proto-X chromosome, one of two 'proto' sex chromosomes evolved from the ancestral autosome pair according to the Ohno's theory (Ohno, 1967). All mammals share this evolutionary domain (placental and not).
- **X-added region (XAR)** including the short arm and the PAR2, which established on X chromosome by translocation from a second autosome. This region is exclusively present in placental mammals.

The gene density of the X chromosome is among the lowest in the genome (7.1 gene per Mb) (Ross et al. 2005). This unusually low gene density is probably a consequence of the

massive expansion of non-coding intergenic sequences that during evolution have been interposing between genes (Bellott et al., 2010).

2.4.2. *X-linked genes*

A total of 1551 genes have been hitherto annotated in the genomic databases (www.ensembl.org/biomart) and, of these, 800 are protein-coding genes. With the more accurate assembly and consequent recalibration of the human X chromosome's gene content, Mueller et al. tested Ohno's law - which states that the gene content of X chromosomes is conserved among placental mammals- by systematically comparing the gene contents of the human and mouse X chromosomes. They found that 18% (144/800) protein-coding genes violate Ohno's law, since they were not shared by the two species. The majority of them (76/144; 52.7%) were acquired independently on the X chromosome since the two lineages began to diverge from a common ancestor 80 million years ago; such independent acquisition apparently occurred through transposition or retroposition from autosomes, or having arisen *de novo*. Among the independently acquired X-linked genes, approximately two-thirds (48/76) are ampliconic (i.e. embedded in duplicated segments of >10 kb in length and exhibiting >99% nucleotide identity), whereas the remaining are multicopy (only the gene structure is duplicated) or single copy genes. Interestingly, ampliconic genes are predicted to have a function in male fitness. Overall, only 31% of the human X-ampliconic genes had orthologs in the other species.

Mueller et al (2013) also reported that most independently acquired human and mouse X-linked genes exhibit high expression in the testis and little or no expression in other tissues. In mice, this prevalent testis expression is related to the male germ cell-restricted expression of these genes regardless of whether they are single, multi-copy or ampliconic (Mueller et al 2013). These novel findings are in line with previous genomic studies reporting an enrichment on the mammalian X chromosomes, compared to the autosomes, for male-specific single and multi-copy genes showing testis-restricted or predominant expression (Wang et al., 2001; Lercher et al., 2003; Mueller et al., 2009; Zheng et al., 2010). Given that the independently acquired genes are expressed predominantly in spermatogenic cells, one might anticipate that loss-of-function mutations affecting these genes or gene families would perturb male gametogenesis.

Given that the X chromosome is enriched with single copy genes expressed during the early stages of murine spermatogenesis, it was originally suggested that mainly pre-meiotic genes were located on the X chromosome (Wang et al 2001). Accordingly, X chromosome is transcriptionally active only in mitotically dividing spermatogonia and in the early meiotic

(pre-pachytene) spermatocytes. During meiosis X-linked genes undergo the so-called meiotic sex chromosome inactivation (MSCI) and thus are transcriptionally silenced (Zheng et al., 2010). However, evidence shows that many microRNAs are expressed also at the pachytene stage, when MSCI occurs, suggesting that a transcriptional activity co-exists also during and after meiosis (Song et al., 2009). The escape from MSCI silencing by X-linked mRNA suggests that they may contribute to MSCI or be involved in post-transcriptional regulation of autosomal mRNA during meiotic and post-meiotic stages of spermatogenesis. In addition, a post-meiotic transcription reactivation has been reported for several multi-copy mouse X-linked gene families (Wang et al., 2005; Mueller et al., 2008) showing higher expression levels compared to single copy genes (Fig. 2.14). It was therefore hypothesized that increasing copy number may be a mechanism to counteract transcriptional repression of the X chromosome in post-meiotic germ cells.

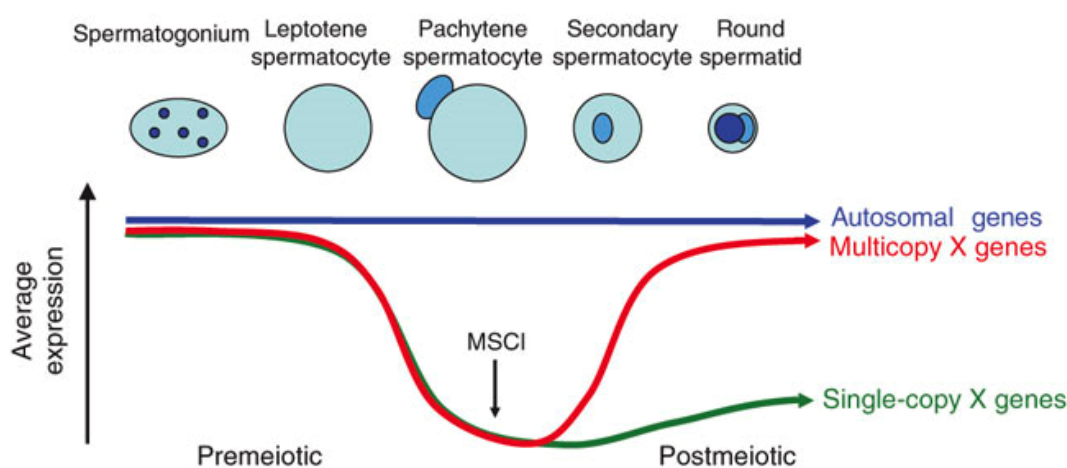


Figure 2.14. Multi-copy genes evade the effects of X chromosome post-meiotic repression in the mouse. Single-copy and multi-copy X-linked genes exhibit similar average levels of expression during pre-meiotic spermatogenesis. All X-linked genes are subsequently silenced during MSCI. Following MSCI, single-copy X-linked genes exhibit low reactivation levels whereas multi-copy X-linked genes exhibit expression levels similar to autosomal genes, thus evading the effects of post-meiotic repression. Figure from Mueller et al. 2008.

The most represented X-linked testis specific gene families are the Cancer Testis (CT) genes which have been suggested to account for 10% of human X-chromosome gene content (Ross et al 2005). CT genes are defined by a unique expression pattern: amongst normal tissues, they are expressed exclusively or predominantly in male germ cells and in embryonic trophoblasts, but their gene products are also found in a significant number of human tumors of different histological origin. At least 70 families of CT genes with over 140 members have been identified so far and recently listed in a database established by the Ludwig Institute for Cancer Research ([http:// www.cta.lncc.br/](http://www.cta.lncc.br/)) (Almeida et al., 2009). The

X-linked CT genes (X-CT) represent more than half of all CT genes and often constitute multi-copy gene families organized in well-defined clusters along the X chromosome, where the different members are arranged into complex direct and inverted repeats (segmental duplications) (Fratta et al., 2011). This feature account for the susceptibility of CT genes to CNVs even though their multi-copy gene status may be a strategy to increase the chance to escape MSCI during meiosis, as observed for mouse X-linked multi-copy genes. The *MAGE* (Melanoma antigen) and *GAGE* (G antigen) are the largest and best-known X-CT gene families containing at least 24 and 16 members, respectively (Stouffs et al., 2009). The biological function of most X-CT genes is still largely unknown. However, evidence is emerging that the best studied of these, the *MAGE* genes, can act as signal transducing transcriptional modulators. Moreover, *MAGE* genes appear to be able to mediate proliferative signals (Park and Lee, 2002; Duan et al., 2003; Glynn et al., 2004). In normal testis, X-CT genes are expressed primarily in the spermatogonia. According to the so-called Rice's theory, such enrichment of male-specific genes on the X chromosome would be related to the accumulation of recessive alleles/genes with beneficial effect for men (*masculinization* of the X chromosome). Indeed, recessive alleles that are beneficial to males will expectedly become fixed more rapidly on the X chromosome than on an autosome (Hurst, 2001) and if these alleles were detrimental to females, their expression could become restricted to male tissues.

Why studying the X chromosome?

Being the "male" chromosome, the Y chromosome has been for decades the main focus of most of the research related to the genetics of male infertility. However, the constant discoveries that throughout time allowed the fine characterization of the sequence and gene content of the X chromosome encouraged researchers to expand their investigation to this chromosome as well.

Two main features make the X chromosome an undeniably attractive object for the study of male infertility. As thoroughly explained in paragraph 2.4.2, this chromosome is full of genes specifically expressed in the testis, thus potentially involved in spermatogenesis. Moreover, with the exception of PAR-linked genes, men are hemizygous for most of the genes located on this chromosome and any *de novo* mutation might have an immediate impact, since no compensation is exerted by another normal allele. Considering that deleterious mutations in crucial spermatogenesis genes cannot be transmitted to future generations, it is highly probable that they arise *de novo* and at a low frequency. For this reason, also private mutations - found only in one infertile patient- might cause infertility. Furthermore,

considering the low prevalence of single gene mutations in candidate spermatogenesis genes, it is currently postulated that infertility should be regarded as a polygenic disease (Cram et al., 2004). In this view, the classical candidate gene approach, focusing on single genes of interest, is a definitely inefficient strategy as shown by the paucity of mutations hitherto identified in the seven X-linked candidate genes studied so far (*AR*, *SOX3*, *USP26*, *NXF2*, *TAF7L*, *FATE* e *AKAP*); for instance, potentially causative mutations have been reported only in the *AR* gene.

As mentioned in paragraph 2.1.2, discovery research has now shifted to whole-genome approaches. High-throughput technologies such as microarrays, including SNP arrays and a-CGH, and next-generation sequencing (NGS) provide the coverage necessary to identify new genetic associations and allows the simultaneous screening of a large number of carefully phenotyped samples, which is a very important requirement for the successful identification of novel genetic associations with infertility. While the application of NGS approaches to male infertility is still dawning and literature is still very poor in this regard, microarrays have already been successfully employed in the last years for the study of CNVs (Tüttelmann et al., 2011; Krausz et al., 2012a; Stouffs et al., 2012; Lopes et al., 2013) and allowed the identification of novel genetic factors, including a number of X-linked CNVs of potential clinical relevance in the etiology of male infertility.

3. AIM OF THE THESIS

The general objective of this thesis was to investigate on the genetic and epigenetic factors potentially involved in idiopathic male infertility. For this purpose, a thorough research was performed in order to determine the role of sex chromosomes-linked CNVs and a high-resolution methylation microarray was employed in order to provide a comprehensive overview of the “normal” human sperm methylome.

The first part of this thesis was aimed to the analysis of not yet identified genetic factors related to the X chromosome, by addressing the following specific issues:

1. To provide an X chromosome-specific outline of CNVs mapping to the genome of infertile patients and to evaluate their potential association with male infertility.
2. To investigate the role of three recurrent X-linked deletions in the etiology of spermatogenic failure.
3. To investigate the role of five selected X-linked duplications in the etiology of spermatogenic failure.

The second part of this thesis was dedicated to the analysis of genetic factors related to the Y chromosome and addressed the following specific issue:

- To investigate whether AZF microdeletions on the Y chromosome are associated with *SHOX* haploinsufficiency.

The third, and last, part of this thesis regards the study of male infertility from the epigenetic standpoint, which provides a comprehensive description of the methylation profile of spermatozoa from normozoospermic subjects; furthermore, the following specific questions were addressed:

- Is there a difference in the DNA methylation pattern between quality-fractioned sperm populations deriving from the same individual?
- Is there an inter-individual variability between the methylation profiles of whole sperm populations and quality-fractioned sperm subpopulations deriving from different normozoospermic subjects?

4. RESULTS

Considering the aforementioned objectives, the results presented in this thesis can be divided accordingly.

To fulfill the objectives related to the first part of this thesis, the initial step consisted in the application of an innovative approach based on a high resolution array- CGH platform specific for the X chromosome, which provided the first detailed analysis of X-linked losses and gains in several hundred subjects with known sperm parameters. This study led to the identification of 73 CNVs (29 losses and 44 gains) - detected in men with both abnormal and normal spermatogenesis – and to the finding that infertile patients with impaired spermatogenesis have a significantly higher burden of CNVs compared to normozoospermic controls (Krausz et al., 2012a). These preliminary data served then for the analysis of selected patient-specific CNVs with potential clinical interest in larger case-control settings. Basically, two separate studies were subsequently performed in order to define the clinical implication of selected X-linked losses and gains. The former study allowed the identification and characterization of three recurrent CNVs, exclusively (CNV67) or predominantly (CNV64, CNV69) found in patients, providing the first evidence of a significant association between recurrent X-linked deletion and spermatogenic failure (Lo Giacco, Chianese et al. 2013). Similarly, the latter study allowed the identification of novel spermatogenesis candidate genes linked to the five selected gains and the discovery of the first recurrent, X-linked gain with potential clinical relevance (Chianese et al., 2014).

With respect to the second purpose of this thesis, the *SHOX* copy number status was analyzed in a large collection of men carrying all type of AZF deletions, including partial deletions and giving special focus to Y-chromosome microdeletions carriers with a normal karyotype. Results from this study showed that both partial and complete Y-chromosome microdeletions in men with 46,XY karyotype are unlikely associated with *SHOX* haploinsufficiency, since of 177 carriers none had *SHOX* deletions.

Finally, to accomplish the third purpose of this thesis, the high-resolution Infinium 450K methylation array was used to obtain the sperm DNA methylation profile at the towering number of 487,517 CpGs sites in a group of eight normozoospermic subjects, the largest number of subjects ever considered by that time. Data from this study, on one hand, allowed defining that the DNA methylation profile is highly conserved among

normozoospermic subjects; on the other hand, the examination of different quality-fractioned sperm populations deriving from the same individual also demonstrated the stability in the sperm DNA methylation pattern (Krausz et al., 2012).

The results briefly resumed above will be presented in detail in the following published articles:

OBJECTIVE 1.

- 1 **High resolution X chromosome-specific array-CGH detects new CNVs in infertile males.** Krausz C, Giachini C, Lo Giacco D, Daguin F, **Chianese C**, Ars E, Ruiz-Castane E, Forti G, Rossi E. *PLoS One*. 2012 7(10):e44887.
- 2 **Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility .** Lo Giacco D and **Chianese C**, Ars E, Ruiz-Castañé E, Forti G, Krausz C. *J Med Genet*. 2014 May;51(5):340-4.
- 3 **X chromosome-linked CNVs in male infertility: discovery of overall duplication load and recurrent, patient-specific gains with potential clinical relevance.** **Chianese C**, Gunning AC, Giachini C, Daguin F, Balercia G, Ars E, Lo Giacco D, Ruiz-Castañé E, Forti G, Krausz C. *PLoS One*. 2014 Jun 10;9(6):e97746.

OBJECTIVE 2.

- **Y-chromosome microdeletions are not associated with SHOX haploinsufficiency.** **Chianese C**, Lo Giacco D, Tüttelmann F, Ferlin A, Ntostis P, Vinci S, Balercia G, Ars E, Ruiz-Castañé E, Giglio S, Forti G, Kliesch S, Krausz C. *Hum Reprod*. 2013 Nov;28(11):3155-60.

OBJECTIVE 3.

- **Novel insights into DNA methylation features in spermatozoa: stability and peculiarities.** Krausz C, Sandoval J, Sayols S, **Chianese C**, Giachini C, Heyn H, Esteller M. *PLoS One*. 2012;7(10):e44479.

4.1. PAPER 1.

High-resolution X chromosome-specific array-CGH detects new CNVs in infertile males.

Krausz C, Giachini C, Lo Giacco D, Daguin F, **Chianese C**, Ars E, Ruiz-Castane E, Forti G, Rossi E. *PLoS One*. 2012 7(10):e44887

High Resolution X Chromosome-Specific Array-CGH Detects New CNVs in Infertile Males

Csilla Krausz^{1,2*}, Claudia Giachini¹, Deborah Lo Giacco^{2,3}, Fabrice Daguin¹, Chiara Chianese¹, Elisabet Ars³, Eduard Ruiz-Castane², Gianni Forti⁴, Elena Rossi⁵

1 Unit of Sexual Medicine and Andrology, Molecular Genetic Laboratory, Department of Clinical Physiopathology, University of Florence, Florence, Italy, **2** Andrology Service, Fundació Puigvert, Barcelona, Spain, **3** Molecular Biology Laboratory, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain, **4** Endocrinology Unit, Department of Clinical Physiopathology, University of Florence, Florence, Italy, **5** Biology and Medical Genetics, University of Pavia, Pavia, Italy

Abstract

Context: The role of CNVs in male infertility is poorly defined, and only those linked to the Y chromosome have been the object of extensive research. Although it has been predicted that the X chromosome is also enriched in spermatogenesis genes, no clinically relevant gene mutations have been identified so far.

Objectives: In order to advance our understanding of the role of X-linked genetic factors in male infertility, we applied high resolution X chromosome specific array-CGH in 199 men with different sperm count followed by the analysis of selected, patient-specific deletions in large groups of cases and normozoospermic controls.

Results: We identified 73 CNVs, among which 55 are novel, providing the largest collection of X-linked CNVs in relation to spermatogenesis. We found 12 patient-specific deletions with potential clinical implication. Cancer Testis Antigen gene family members were the most frequently affected genes, and represent new genetic targets in relationship with altered spermatogenesis. One of the most relevant findings of our study is the significantly higher global burden of deletions in patients compared to controls due to an excessive rate of deletions/person (0.57 versus 0.21, respectively; $p = 8.785 \times 10^{-6}$) and to a higher mean sequence loss/person (11.79 Kb and 8.13 Kb, respectively; $p = 3.435 \times 10^{-4}$).

Conclusions: By the analysis of the X chromosome at the highest resolution available to date, in a large group of subjects with known sperm count we observed a deletion burden in relation to spermatogenic impairment and the lack of highly recurrent deletions on the X chromosome. We identified a number of potentially important patient-specific CNVs and candidate spermatogenesis genes, which represent novel targets for future investigations.

Citation: Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, et al. (2012) High Resolution X Chromosome-Specific Array-CGH Detects New CNVs in Infertile Males. PLoS ONE 7(10): e44887. doi:10.1371/journal.pone.0044887

Editor: Laszlo Orban, Temasek Life Sciences Laboratory, Singapore

Received: February 29, 2012; **Accepted:** August 15, 2012; **Published:** October 9, 2012

Copyright: © 2012 Krausz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Tuscan Regional Health Research Program ("Progetto Salute 2009") to GF, the Italian Ministry of University (grant PRIN 2010-2012 to CK), the Italian Telethon (Grant GGP08204 to CK) and the Spanish Ministry of Health (grant FIS-08/1138, grant FIS-11/02254). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: c.krausz@dfc.unifi.it

Introduction

Male factor infertility affects about 7% of men in the general population and the etiology of altered spermatogenesis remains unknown in about 40% of cases ("idiopathic infertility") and it is likely that a large proportion of them are caused by still unknown genetic factors [1]. Nevertheless, besides abnormal karyotype and Y chromosome microdeletions no other recurrent genetic anomalies have been identified in men with primary testicular failure, raising questions about the appropriateness of the investigative approaches used so far [2–4]. The first innovative study applying whole-genome analysis of SNPs and the successive follow-up study failed in leading to the identification of recurrent genetic factors with large effect size [5,6]. Recently, high resolution array Comparative Genomic Hybridisation (array-CGH) studies identified new spermatogenesis candidate genes on autosomes and on the X chromosome and some recurring and private patient-specific CNVs with potential clinical interest [7,8].

Both sex chromosomes are enriched with genes prevalently or exclusively expressed in the testis [9,10]. Nevertheless, only Y chromosome-linked Copy Number Variants (CNVs) and Y-linked genes have been demonstrated as important contributors to impaired sperm production in humans [for review see [11,12]]. In particular, the so called AZoospermia Factor (AZF) regions on the Yq have been found deleted in about 5–10% of azoospermic men (absence of spermatozoa in the ejaculate) and 2–5% of severe oligozoospermic men (<5 millions spermatozoa in the ejaculate). Data on the potential role of X-linked gene products in spermatogenesis derive mainly from model organisms and a higher than expected number of X-linked spermatogenesis genes have been identified [10,13]. The apparent paucity of information in humans is probably related to the scarcity of X-linked genes studied (only eight), none of which yet described as causative, except for the *AR* gene [14]. Similarly, the question whether the X chromosome contains AZF-like regions has not been sufficiently explored so far.

In order to advance the understanding of the role of X-linked CNVs and genes in male infertility, we applied an innovative approach based on high resolution X chromosome specific array-CGH. Given that such a detailed analysis of the X chromosome has not been published until now and the testicular function of subjects included in the Genomic Variant Database is unknown (except for 30 X-linked CNVs (23 duplications and 7 deletions) reported in the recent paper by Tuttelmann et al. [7]), ours is the first study providing a detailed analysis of X-linked losses and gains in several hundred subjects with known sperm parameters.

Materials and Methods

Subjects

The local Ethical Committees of the University Hospital Careggi and the Fundació Puigvert approved the study. All participants signed an informed consent. We analyzed with array-CGH 96 idiopathic infertile subjects with different grade of spermatogenic impairment (49 azoospermic, 25 cryptozoospermic and 22 oligozoospermic men) and 103 normozoospermic men. Infertile patients were selected on the basis of a comprehensive andrological examination including medical history, semen analysis, scrotal ultrasound, hormone analysis, karyotype and Y chromosome microdeletion screening. Patients with mono- or bilateral cryptorchidism, varicocele grades 2 and 3, obstructive azoospermia, recurrent infections, iatrogenic infertility, hypogonadotrophic hypogonadism, karyotype anomalies, Y chromosome microdeletions including partial deletions of the *AZF* region, and partial *AZF* duplications and patients with non-Italian or non-Spanish origin were excluded. Testis histology was available for 47 men. Controls in the Spanish cohort were fertile normozoospermic men undergoing pre-vasectomy, whereas the Italian control cohort included normozoospermic volunteers not belonging to infertile couples (60% with proven fertility). The ethnic/geographic composition was similar in the control and patient groups (40% Spanish and 60% Italians). In the second part of the study, we performed a case-control association study reaching a total of 359 patients and 370 normozoospermic controls on 13 selected CNVs which appeared to be specific to infertile men based on the array-CGH analysis. Detailed phenotypic data relative to the study populations are provided in Table 1.

Methods

Germline DNA was extracted from peripheral blood samples in all the participants with standard methods.

Array-CGH. Customized array-CGH platforms (custom 8×60 K, Agilent Technologies, Santa Clara, CA, USA) were generated using the eArray software (<http://earray.chem.agilent.com/>); 53069 probes (60-mer oligonucleotides) were selected from those available in the Agilent database and cover the whole chromosome X, including Xp and Xq pseudoregions, with a medium resolution of 4 Kb. Four replicate probe groups, with every probe present in two copies on the platform, were designed in regions containing mouse infertility-associated genes i.e. sperm protein associated with the nucleus, X-linked family members (*SPANX*); testis expressed 11 *TEX11*, TAF7-like RNA polymerase II, TATA box binding protein (TBP)-associated factor (*TAF7L*) and). In these regions, the medium resolution is 2 Kb. The array also included, for the normalization of copy number changes, Agilent control clones spread along all autosomes (6842 probes). As a reference DNA, we used the same normozoospermic subject for all the study population. This control DNA was already characterized for CNV content in previous array-CGH experiments against eight different normospermic controls and presented

one private gain of 27 Kb mapping to Xcentr which was not considered for the frequency analyses. 300 ng of test DNA and control DNA were double-digested with RsaI and AluI (Promega) for 1 hour at 37°C. After digestion, samples were incubated at 65°C for 20 minutes to inactivate the enzymes, and then labeled by random priming (Agilent Technologies) for 2 hours using Cy5-dUTP for the test DNA and Cy3-dUTP (Agilent Technologies) for the control DNA. Labeled DNAs were incubated at 65°C for 10 minutes and then purified with Microcon YM-30 filter units (Millipore, Billerica, USA). Every purified sample was brought to a total volume of 9.5 µl in 1×TE (pH 8.0, Promega), and yield and specific activity were determined for each sample using a NanoDrop ND-1000 UV-VIS Spectrophotometer (Labtech International LTD). The appropriate cyanine 5- and cyanine 3-labeled samples were combined in a total volume of 16 µl. After sample denaturation and pre-annealing with 5 µl of Human Cot-1 DNA (Invitrogen, Carlsbad, CA), hybridization was performed at 65°C with shaking for 24 hours. After two washing steps, the array was analyzed through the Agilent scanner and the Feature Extraction software (v10 1.1.1). Graphical overview was obtained using the DNA Analytics (v4.0.73). All the array experiments were analysed using the ADM-2 algorithm at threshold 5. Aberrant signals including 3 or more adjacent probes were considered as genomic CNVs (Figure S1). The positions of oligomers refer to the Human Genome March 2006 assembly (hg18). All experimental data was submitted to GEO repository with the following Series accession number: GSE37948.

Molecular genetic analyses for confirmation of array-CGH data and for the case-control study

Molecular analysis of deletions. For the first step screening as for the confirmatory step, we performed PCR protocol in a final volume of 10 µl containing 70 ng of genomic DNA, 3 mM MgCl₂, 400 µM deoxynucleotides triphosphates, 10 pmol of specific primers, 50 U/ml of Taq DNA Polymerase (Promega PCR MASTER MIX 2X). All the primers for the first step screening had an optimal annealing temperature between 58–60°C and suspected deletions were further confirmed by i) lowering the annealing temperature (55°C); ii) performing additional PCRs with alternative primers (see details in the Table S1).

Molecular analysis of gains and the loss CNV31. Gains and loss CNV31 screening were performed using pre-designed TaqMan® Copy Number Assays or Custom TaqMan® Copy Number Assays (Applied Biosystems). All assays were conducted using three or four replicates for each sample (on the basis of the assay quality), in a final volume of 20 µl according to the manufacturer's instructions. The reaction mix components were: 1X TaqMan® Genotyping Master Mix, 1X TaqMan® Copy Number Assay, 1X TaqMan® Copy Number Reference Assays, 10 ng of genomic DNA. Briefly, the TaqMan® Copy Number Assay – containing two specific primers and a FAMTM dye-labeled MGB probe to detect the genomic DNA target sequence – is run in duplex with the TaqMan® Copy Number Reference Assays – containing two primers and a VIC® dye-labeled TAMRATM probe to detect the genomic DNA reference sequence. On each plate the same normozoospermic control used as reference DNA for array-CGH experiments (calibrator sample), the DNA sample of the CNV carrier and the No Template Control (NTC) were run. The CopyCaller Software™ was used for post-PCR data analysis for all the copy number quantitation experiments. Information about qPCR probes are provided in Table S2.

Table 1. Clinical description of the study population.

A				
SPERM COUNT		PATIENTS (n = 359)	CONTROLS (n = 370)	
Total sperm count (10 ⁶)	median (25 th –75 th percentile)	2.6 (0.00–13.62)	263.20 (159.00–405.50)	
	mean ± SD	8.77±12.72	311.79±199.99	
Sperm concentration (10 ⁶ /ml)	median (25 th –75 th percentile)	0.90 (0.00–4.40)	76 (50.00–117.50)	
	mean ± SD	2.56±3.27	91.32±59.64	
B				
GEOGRAPHIC ORIGIN	SEMEN PHENOTYPE	ARRAY-CGH STUDY (n = 96)	CASE-CONTROL STUDY (n = 263)	TOTAL (n = 359)
ITALIAN (n = 233)	azoospermic	15	42	57
	cryptozoospermic	20	17	37
	oligozoospermic	22	117	139
	severe oligozoospermic (<5×10 ⁶ /ml)	17	47	-
	moderate oligozoospermic(<20×10 ⁶ /ml)	0	12	-
SPANISH (n = 126)	azoospermic	34	15	49
	cryptozoospermic	5	36	41
	oligozoospermic	0	36	36
	severe oligozoospermic (<5×10 ⁶ /ml)	0	24	-
	moderate oligozoospermic(<20×10 ⁶ /ml)	0	12	-
C				
HORMONAL PARAMETERS	ARRAY-CGH STUDY (n = 96)	CASE-CONTROL STUDY (n = 263)	TOTAL (n = 359)	
FSH (U/L)				
Mean ± SD	13.40±9.09	11.13±10.08	11.74±9.84	
Median (25 th –75 th percentile)	11.38 (5.30–19.0)	8.10 (4.27–14.30)	8.70 (4.40–15.0)	
LH (U/L)				
Mean ± SD	5.02±2.34	5.46±4.02	5.36±3.72	
Median (25 th –75 th percentile)	4.45 (3.30–6.21)	4.60 (3.20–6.75)	4.60 (3.20–6.70)	
TESTOSTERONE (ng/ml)				
Mean ± SD	5.18±2.94	4.49±2.08	4.64±2.30	
Median (25 th –75 th percentile)	4.80 (3.51–6.40)	4.20 (3.40–5.20)	4.20 (3.40–5.40)	
TESTIS VOLUME (ml)				
Mean ± SD	11.55±4.55	12.89±4.32	12.54±4.42	
Median (25 th –75 th percentile)	11.50 (8.0–14.0)	13.0 (9.0–14.87)	13.0 (9.0–14.50)	

A) Semen phenotype of the entire study population (array-CGH and case-control study); B) Description of all analyzed patients (array-CGH and case-control study) according to their geographic origin and semen phenotype; C) Hormonal levels and testis volumes of all analyzed patients (array-CGH and case-control study). doi:10.1371/journal.pone.0044887.t001

Statistical analysis

Statistical analyses were performed using the statistical package SPSS (version 17.0.1, Chicago, IL, USA). Non-parametric Mann-Whitney U test was performed for comparisons of: i) median values of CNV number and DNA change between patients and controls; ii) median values of sperm concentration and total sperm count in relationship with CNV number. Frequencies were compared by Fisher exact test.

Results

Characterization of X-chromosome linked CNVs

We performed a high resolution array-CGH analysis using a microarray containing probes densely covering the complete human X chromosome (average resolution: 4 kb). Of the 199 subjects analyzed (96 idiopathic infertile subjects and 103

normozoospermic men), 97 (36 patients and 61 controls) showed the lack of CNVs, whereas the remaining 102 samples were found to carry 73 CNVs (44 gains and 29 losses) (Tables 2, 3, and 4). Thirty-two CNVs intersected genes/transcription units based on data available in genomic databases. As shown in Figure 1, CNVs were evenly distributed along the X chromosome with higher density in the PAR1.

Since homologous sequences at the border of a CNV may act as a substrate for non-allelic homologous recombination (NAHR), we checked the nature of regions flanking (between the minimum-maximum size of the CNV and approximately up to 1 Mb from the maximum size) the identified CNVs in order to understand whether NAHR is likely to occur (UCSC Genome Browser). Highly homologous sequences were identified only in 19% of CNVs, indicating that NAHR is not involved in the majority of observed CNVs. This figure was concordant with other observa-

Table 2. List of the 31 patient-specific (not found in normozoospermic controls) CNVs detected by array-CGH and their description according to type, gene location (NO = no gene found within) and occurrence in the Database of Genomic Variants (DGV).

CNV type	CNV code	Region	Size (Kb)	Start position	End position	Coding sequences within the CNV*	DGV	Frequency
LOSSES	17	Xp22.31	31.70	6,594,834	6,626,533	NO		1
	18	Xp22.31	82.00	6,756,310	6,838,310	NO	Variation_8908 Variation_53018 Variation_34619	1
	22	Xp22.11	24.35	22,969,648	22,993,957	NO		1
	23	Xp21.3	6.69	25,274,024	25,280,712	NO		1
	24	Xp21.3	67.33	26,891,769	26,959,101	NO		1
GAINS	31	Xp11.23	81.13	47,766,391	47,847,516	ZNF630	Variation_96640 Variation_9861 Variation_53005	1
	32	Xp11.22	4.15	52,065,798	52,069,943	NO		1
	50	Xq22.1	45.36	101,803,578	101,848,935	ARMX5-GPRASP2		1
	54	Xq24	44.85	118,281,024	118,325,874	NO		1
	56	Xq25	86.07	124,632,886	124,718,959	NO		1
	57	Xq25	188.03	124,929,673	125,117,699	NO	Variation_52924	1
	61	Xq27.2	4.67	140,773,893	140,778,561	MAGEC3		1
	66	Xq27.3	7.35	145,030,566	145,037,917	NO		1
	67	Xq28	5.42	148,456,474	148,461,889	NO		1
	71	Xq28 (PAR)	290.99	154,586,913	154,877,901	SPRY3 VAMP7		1
GAINS	1a	Xp22.33 (PAR)	224.83	1,544	226,372	PLCXD1 GTPBP6 PPP2R3B		1
	5	Xp22.33 (PAR)	160.10	302,644	462,740	NO		1
	11	Xp22.33 (PAR)	39.73	1,347,599	1,387,328	CSF2RA		1
	14	Xp22.33 (PAR)	1.40	1,896,197	1,897,608	NO	Variation_31542	2
	19	Xp22.31	245.03	7,002,649	7,247,676	STS HDHD1A		1
	20	Xp22.2	602.10	11,104,518	11,706,614	ARHGAP6 AMELX MSL3		1
	25	Xp21.3	60.40	27,277,529	27,337,933	NO		1
	26	Xp21.1	88.57	37,168,387	37,256,960	PRRG1		1
	27	Xp21.1	9.61	37,242,364	37,251,969	NO		1
	30	Xp11.3	81.13	47,766,391	47,847,516	ZNF630	Variation_53003 Variation_9343 Variation_83491	1
	38	Xq13.2	8.98	72,202,996	72,211,976	NO		1
	39	Xq21.1	21.98	76,992,067	77,014,050	MAGT1		1
	40	Xq21.1	5.28	80,112,246	80,117,526	NO	Variation_83611	2
	55	Xq24	206.90	118,691,020	118,897,917	SEPT6 ANKRD58 RPL39 SNORA69 UPF38 RNFT13A NDUFA1		1
	58	Xq25	9.68	125,143,278	125,152,957	NO		1
	60	Xq26.3	42.30	134,585,636	134,627,936	NO	Variation_52936	1

*CNV minimum size.
doi:10.1371/journal.pone.0044887.t002

Table 3. List of the 33 control-specific (not found in idiopathic patients) CNVs detected by array-CGH and their description according to type, gene location (NO=no gene found within) and occurrence in the Database of Genomic Variants (DGV).

CNV type	CNV code	Region	Size (Kb)	Start position	End position	Coding sequences within the CNV*	DGV	Frequency
LOSSES	5.B	Xp22.33	12.63	701,071	713,696	NO		1
	25.A	Xp21.2	9.69	31,282,923	31,292,613	DMD		1
	25.B	Xp21.1	28.26	33,953,232	33,981,492	NO	Variation_7783	2
	33.A	Xp11.21	58.89	56,403,390	56,462,278	NO		1
	53.A	Xq24	170.73	118,278,913	118,449,646	SLC25A43		1
	58.A	Xq25	12.68	125,198,109	125,210,792	NO		1
	60.A	Xq26.3	50.84	134,801,361	134,852,198	SAGE1		1
	60.D	Xq27.1	217.83	140,175,103	140,392,930	NO		1
	66.A	Xq28	37.12	147,393,583	147,430,698	AFF2		1
	71.A	Xq28 (PAR)	122.36	154,755,542	154,877,901	VAMP7		1
GAINS	4.A	Xp22.33 (PAR)	237.08	153,373	390,452	PLCXD1 GTPBP6 PPP2R3B		1
	5.A	Xp22.33 (PAR)	241.98	674,222	916,206	NO		1
	5.C	Xp22.33 (PAR)	420.72	747,358	1,168,080	NO		1
	12.A	Xp22.33 (PAR)	6.61	1,693,897	1,700,511	ASMT		1
	12.B	Xp22.33 (PAR)	683.74	1,716,023	2,399,766	ASMT DHRSX		1
	15.A	Xp22.33 (PAR)	27.94	2,382,699	2,410,643	DHRSX	Variation_83270	1
	15.B	Xp22.33/22.32	280.09	4,206,493	4,486,580	NO		1
	16.A	Xp22.31	1609.42	6,487,238	8,096,662	HDHD1 STS VCX PNPLA MIR651		1
	19.A	Xp22.31	129.96	7,961,788	8,091,751	MIR651	Variation_9337	1
	19.B	Xp22.31	177.54	8,411,159	8,588,699	KAL1		1
	20.A	Xp22.2	665.88	14,590,604	15,256,487	GLRA2 FANCB MOSPD2 ASB9 ASB11 PIGA		1
	20.B	Xp22.13	13.34	18,018,894	18,032,238	NO		1
	25.C	Xp21.1	185.02	34,931,807	35,116,827	NO		1
	25.D	Xp21.1	215.00	35,269,628	35,484,626	NO		1
	31.A	Xp11.23	78.87	48,021,982	48,100,848	SSX3		1
	34.A	Xp11.12	48.06	56,870,427	56,918,489	NO		1
	36.A	Xq11	716.03	63,925,948	64,641,977	ZC4H2 ZC3H12B		1
	38.A	Xq13.2	192.04	74,375,875	74,567,915	UPRT ZDHHC15	Variation_74012	1
	38.B	Xq13.3	153.23	75,123,387	75,276,621	NO		1
	55.A	Xq25	53.80	120,385,787	120,439,584	NO		1
	59.A	Xq26.3	24.69	134,151,039	134,175,725	NO		1
	60.B	Xq26.3	13.45	136,050,422	136,063,872	NO		1
	60.C	Xq26.3	91.26	137,089,527	137,180,783	NO		1

*CNV minimum size.

doi:10.1371/journal.pone.0044887.t003

tions reporting a similar frequency of potential NAHR targets [15]. It is interesting to note that in some areas (Xp11.12-q21.1) only duplications were found, whereas from Xq27.1-q27.3 only deletions were detected. One of the PAR1-linked losses (CNV15) was found in 23 patients and only once in controls (Figure 1b). This small CNV has already been described in the Database of Genomic Variants (DGV) both as loss and gain. This CNV was situated inside a 3914 bp Simple Tandem Repeat which included two Segmental Duplications (respectively of 1498 bp and 1444 bp) that therefore may act as substrate for NAHR. This mechanism may have lead also to reciprocal duplication and in fact CNV14,

identified in our study, is the reciprocal duplication of CNV15. No genes were identified inside or nearby CNV14/15 which made it difficult to attribute a pathogenic role to this loss. Moreover, the same sequence was present also on the Y chromosome which further complicated the interpretation of the results.

Considering the size of detected CNVs, which ranged from 1.4 Kb to 1609 Kb (Tables 2, 3, and 4), we noticed that losses were typically of small/medium size and only 17% of them were large (Figure 2). Conversely, large gains represented 48% of the total CNVs and the difference between frequencies of losses and gains of >100 Kb was statistically significant ($p = 0.012$). Small

Table 4. List of CNVs found by array-CGH considering their occurrence in controls and in patients with their description according to type, gene location (NO = no gene found within) and presence in the Database of Genomic Variants (DGV).

CNV code	CNV type	Region	Size (Kb)	Start position	End position	Coding sequences within the CNV*	DGV	Frequency in patients	Frequency in controls
Control-enriched CNVs									
12	GAIN	Xp22.33 (PAR)	17.30	1,693,897	1,711,194	ASMT	Variation_83259	1	3
Patient-enriched CNVs									
15	LOSS	Xp22.33 (PAR)	1.40	1,896,197	1,897,608	NO	Variation_104545	23	1
64	LOSS	Xq27.3	3.92	143,436,347	143,440,268	NO	Variation_115340	8	5
69	LOSS	Xq28	11.77	154,044,877	154,056,645	NO		7	3
Common CNVs									
16	LOSS	Xp22.32	7.76	4,250,413	4,258,174	NO	Variation_52995	2	1
35	GAIN	Xp11.1	117.14	57,318,438	57,435,573	FAAH2		1	1
49	GAIN	Xq22.1	5.30	100,942,190	100,947,490	NO		2	2
51	GAIN	Xq22.2	34.69	103,152,319	103,187,013	H2BFWT H2BFM	Variation_3254	5	4
68	GAIN	Xq28	105.66	148,686,631	148,792,286	MAGEA8	Variation_31571	2	2

*CNV minimum size.

doi:10.1371/journal.pone.0044887.t004

CNVs (<10 Kb) were more frequently found in patients in respect to controls whereas large gains have been found mainly in controls (Figure 2).

According to the Database of Genomic Variants (DGV) website, losses/gains were divided into “known” and “novel”, identifying 21 novel losses and 34 novel gains (Tables 2, 3, and 4). Among the 73 CNVs, 31 (15 losses and 16 gains) were found only in patients, “patient-specific” (Table 2) and 33 (10 losses and 23 gains) were found only in the control group, “control-specific” (Table 3). Of the remaining 9 CNVs, only one gain (CNV12) was found more frequently among controls whereas those resulting more frequent among patients (“patient-enriched”) were deletions. The rest (4 gains and 1 deletion) were found to equally occur in both patients and controls (Table 4). These data suggest that gains are less likely to affect spermatogenesis since 63% of them (28/44) were found also in normozoospermic controls. On the contrary, deletions were less frequent in controls (11/29; 38%) indicating that in the presence of a deletion an abnormal sperm phenotype is more likely to occur. A general outline of the array-CGH findings with phenotypic description is provided in Table S3.

CNV burden

In order to assess the potential impact of CNVs in cases versus controls, we used two primary measures of CNV burden: the mean size and the mean number of CNVs/individual (Table 5A). The mean value of losses bp was significantly higher in patients than in controls (11.79 Kb and 8.13 Kb, respectively; $p = 3.435 \times 10^{-4}$). All losses were confirmed by PCR plus/minus or Real Time PCR, except for PAR-linked losses ($n = 4$), for which no suitable assay could be designed. The number of CNVs/person was significantly higher in patients compared to controls ($p = 0.002$) and depended on the overrepresentation of losses in the former group (0.57 versus 0.21; $p = 8.785 \times 10^{-6}$) (Table 5). CNV15, the most frequently found loss appears to be the major contributor to the deletion burden, however even without this loss the number of losses/person is significantly higher in the patient's group ($p = 0.041$). Phenotypic description of patients (loss-carriers and no CNV-carriers) is provided in Table S4. Although the frequency of patients with more than one CNV ($n = 19$; 19.8%) was nearly twice that of controls ($n = 11$; 10.7%), the difference did not reach statistical significance ($p = 0.078$). On the other hand, comparing the frequencies of subjects with ≥ 1 CNV in cases versus controls, we observed a highly significant difference when considering the total number of CNVs ($p = 0.003$) and of losses ($p < 0.001$) (Table 5B).

CNVs and semen parameters

A significant association with sperm concentration and total sperm number was observed among patients when considering the total CNV number (Table 6). Patients with more than 1 CNV had a significantly lower sperm concentration and total sperm count than those with ≤ 1 CNV ($0.2 \pm 0.6 \times 10^6/\text{ml}$ versus $1.0 \pm 2.0 \times 10^6/\text{ml}$; $p < 0.022$; $2.3 \pm 4.6 \times 10^6$ versus $1.0 \pm 3.3 \times 10^6$; $p < 0.032$). The maximum number of CNVs/subject was three, and of the five patients with three CNVs four were azoospermic and one was severely oligozoospermic with <1 million spermatozoa/ejaculate (Table S6). All of them had at least one private CNV (uniquely found in this patient), and only one patient (07-170) shared two recurrent CNVs with two others (07-13, 07-30). Given that the selection of patients was based on the absence of known causes of spermatogenic failure, subjects with multiple CNVs did not show any additional andrological anomaly or other relevant diseases. Semen parameters and testis histology of patients and controls with >1 CNVs are reported in Table S5, 6.

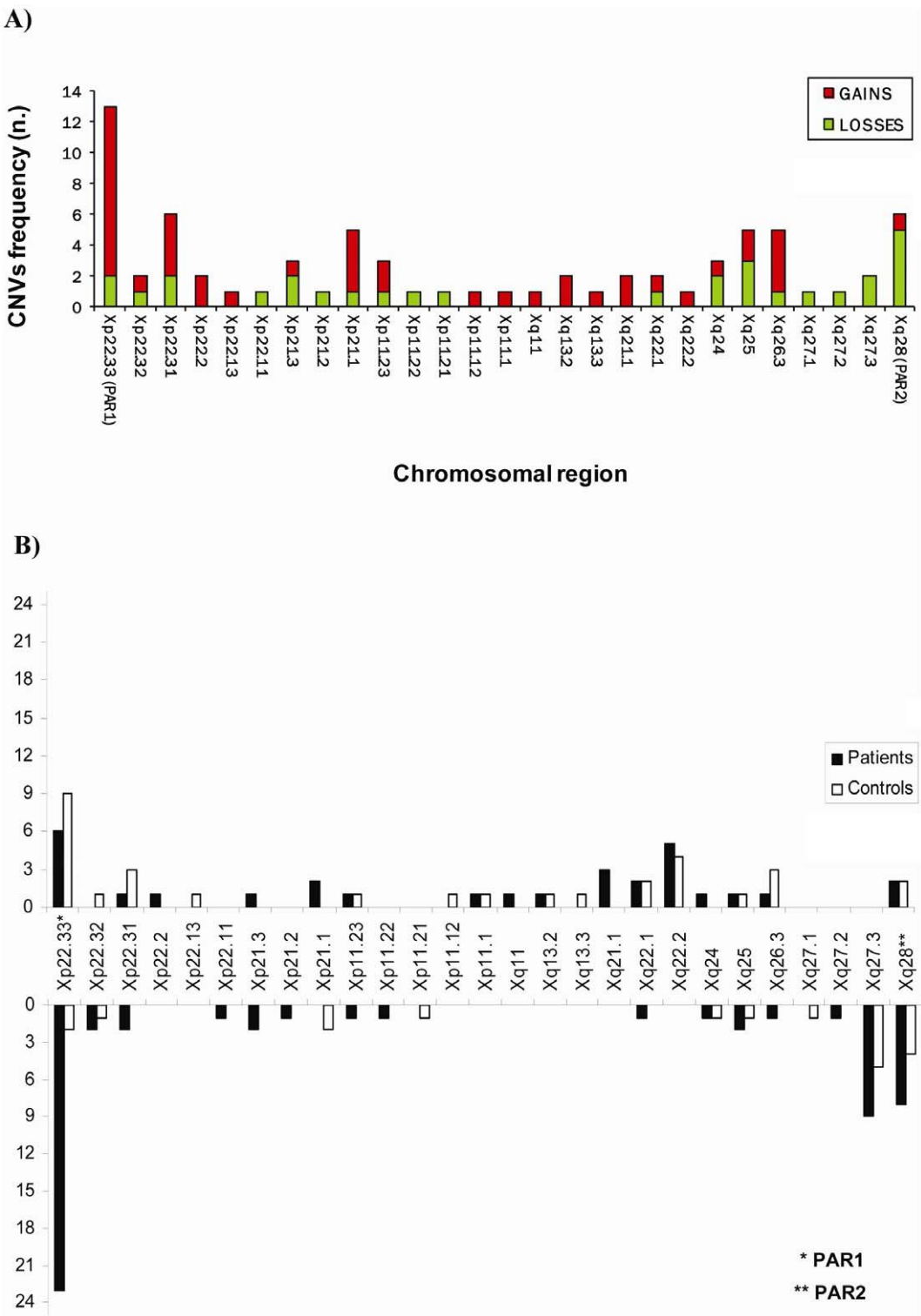
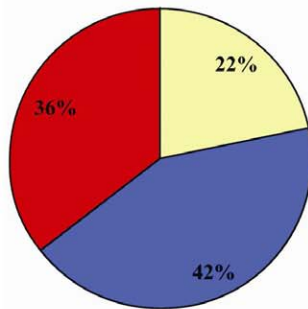


Figure 1. Schematic representation of the distribution of the 73 CNVs (44 gains and 29 losses) along the X chromosome identified by high resolution X chromosome specific array-CGH analysis. A) The histogram shows that the 73 CNVs were evenly distributed along the X chromosome but displayed a higher density in the pseudoautosomal region 1, PAR1 (Xp22.33). B) The frequency of gains (upwards) and losses (downwards) per X chromosome region in patients and controls are indicated.
doi:10.1371/journal.pone.0044887.g001

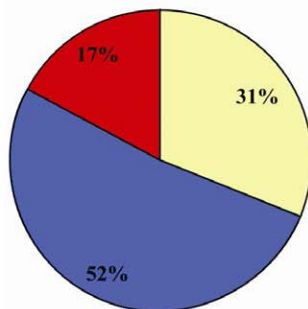
Screening for selected deletions

To further investigate the potential clinical implications of losses, 13 patient-specific deletions were subsequently screened in a

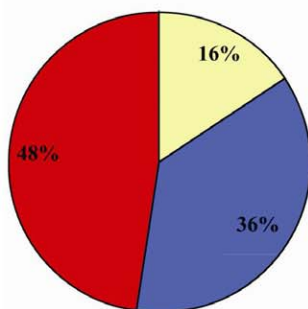
large group of infertile and normozoospermic men: excluding CNV66, they all remained patient-specific (Table 7). Due to the rarity of the 12 patient-specific losses, statistically significant

A. Gains + Losses

A. Gains + Losses (n=73)				
	<10kb	10-100 kb	>100 kb	Novel
Patient-specific	10	14	7	23
Control-specific	2	14	17	29
Patient-enriched	2	1	-	1
Control-enriched	-	1	-	-
Common	2	1	2	2

B. Losses

B. Losses (n=29)				
	<10kb	10-100 kb	>100 kb	Novel
Patient-specific	5	8	2	11
Control-specific	1	6	3	9
Patient-enriched	2	1	-	1
Control-enriched	-	-	-	-
Common	1	-	-	-

C. Gains

C. Gains (n=44)				
	<10kb	10-100 kb	>100 kb	Novel
Patient-specific	5	6	5	12
Control-specific	1	8	14	20
Patient-enriched	-	-	-	-
Control-enriched	-	1	-	-
Common	1	1	2	2



Figure 2. Array-CGH study. distribution of the 73 CNVs according to their size: small (<10 Kb), medium (10–100 Kb) and large (>100 Kb) referred to A) all CNVs (44 gains and 29 losses); B) losses; C) gains. Losses were typically of small/medium size (52%) whereas gains are generally of larger size (48%). On the side, tables display the number of A) all CNVs; B) losses; C) gains of different size and categorized according to their occurrence in patients/controls: i) “patient-specific” when found only in patients; ii) “control-specific” when found only in controls; iii) “patient-enriched” when found predominantly in patients; iv) “control-enriched” when found predominantly in controls; v) “common” when found at a similar frequency in patients and controls.

doi:10.1371/journal.pone.0044887.g002

differences were not observed in their frequencies compared to the control group. In fact, 8/12 were private (found in a single individual) whereas only 4 were recurrent with a still relatively low frequency (0.5–1.1%).

Recurrent patient-specific CNVs. Among the patient-specific recurrent CNVs, three deletions are of major interest. CNV67, observed in 1.1% of patients may remove (considering its maximum size) the melanoma antigen family A, 9B (*MAGEA9B*), which belongs to the Cancer Testis Antigens (CTAs) gene family,

Table 5. Array-CGH study: Comparison between patients and controls of the mean number and mean extension of CNVs (A) as well as the number of all subjects bearing more than one CNV (B).

A	PATIENTS (n = 96)		CONTROLS (n = 103)		p	
	Mean CNV number \pm sd	Mean CNV extension (Kb) \pm sd	Mean CNV number \pm sd	Mean CNV extension (Kb) \pm sd	p1	p2
LOSSES+GAINS	0.87 \pm 0.85	36.21 \pm 85.4	0.54 \pm 0.76	73.87 \pm 222.08	2.095 $\times 10^{-3}$	0.113
LOSSES	0.57 \pm 0.64	11.79 \pm 38.43	0.21 \pm 0.46	8.13 \pm 32.30	8.785 $\times 10^{-6}$	3.435 $\times 10^{-4}$
GAINS	0.30 \pm 0.54	24.42 \pm 76.50	0.33 \pm 0.62	65.74 \pm 220.07	0.862	0.733
B	PATIENTS (n = 96)		CONTROLS (n = 103)		p	
					OR (95% CI)	
	≥ 1 CNV/subject		42		0.003	
	60				1.5 (1.2–2.0)	
	≥ 1 loss/subject		20		<0.001	
	47				2.5 (1.6–3.9)	
	≥ 1 gain/subject		28		0.874	
	25				-	

sd = standard deviation. OR = odds ratio. CI = confidence interval. p1 refers to the mean number of CNV/subject. p2 refers to the mean DNA change/subject.
doi:10.1371/journal.pone.0044887.t005

expressed exclusively in the testis with the highest expression level in spermatocytes and in some tumour cell lines [16]. This deletion may also affect additional genes with prevalent or exclusive expression in the testis such as other CTAs and the following: transmembrane protein 185A (*TMEM185A*), chromosome X open reading frame 40A (*CXorf40A*), X linked heat shock transcription factor family (*HSEF*) all situated at <1 Mb from the deletion. Phenotypes of patients with this deletion ranged from azoospermia due to Sertoli Cell Only Syndrome (SCOS, [17]) to oligozoospermia. CNV 31 presents a reciprocal duplication (CNV30, Table 2) and was observed in 4 patients (two found by array-GH and two by qPCR) and 0/325 controls. CNVs 30/31 affect the dosage of zinc finger protein 630 (*ZNF630*), a gene with unknown function; however, considering their maximum extension, additional genes with exclusive expression in the testis such as the sperm acrosome associated 5 (*SPACA5*,/*SPACA5b*) are also involved. CNV32 does not remove any gene directly, but it is situated within an area abundant in CTA genes. In order to define whether the underlying mechanism of these deletions is NAHR we analyzed the flanking regions. Only CNV 30/31 showed

Segmental Duplications (SD) which may explain the recurrence of deletion/duplication events. Although also CNV67 was found in 4 patients, this deletion does not have a reciprocal duplication and it is not flanked by SDs. An alternative mechanism for the formation of CNV67 could be Non Homologous End Joining (NHEJ), since substrates for this mechanism are highly represented in this area (many LINE and Alu elements). However this hypothesis requires further confirmation by the fine mapping of the breakpoints.

Private patient-specific CNVs. Concerning private patient-specific deletions, which were found only in single patients, we observed two deletions directly affecting gene dosage. CNV50 removes the *ARMCX5-GPRASP2* read-through (*ARMCX5-GPRASP2*) genes for which no testis expression data are available. The carrier of this deletion suffers from azoospermia due to SCOS. CNV61, observed in one azoospermic man, removes another CTA family member, the melanoma antigen family C, 3 (*MAGEC3*). This deletion may also affect other neighbouring CTA genes, such as the melanoma antigen family C, 1 (*MAGEC1*) and Sperm protein associated with the nucleus, X-linked, family

Table 6. Array-CGH study: comparison of patients' semen parameters according to the number of CNVs.

PATIENTS (n = 96)				
	SPERM CONCENTRATION (n $\times 10^6$ /ml)	p	TOTAL SPERM NUMBER (n $\times 10^6$)	p
0 CNV (n = 36)	1.2 \pm 2.4 (0.01; 0.0–12.0)		2.9 \pm 5.7 (0.01; 0.0–30)	
≥ 1 CNV (n = 60)	0.6 \pm 1.3 (0.0; 0.0–6.2)	0.068	1.6 \pm 3.4 (0.0; 0.0–17.4)	0.075
0 LOSS (n = 49)	1.0 \pm 2.1 (0.01; 0.0–12.0)		2.7 \pm 5.0 (0.01; 0.0–30.0)	
≥ 1 LOSS (n = 47)	0.6 \pm 1.5 (0.0; 0.0–6.2)	0.053	1.4 \pm 3.6 (0.0; 0.0–17.4)	0.051
0 GAIN (n = 71)	1.0 \pm 2.1 (0.0; 0.0–12.0)		2.4 \pm 4.9 (0.0; 0.0–30.0)	
≥ 1 GAIN (n = 25)	0.4 \pm 0.7 (0.0; 0.0–2.3)	0.185	1.3 \pm 2.3 (0.0; 0.0–6.4)	0.215
≤ 1 CNV (n = 77)	1.0 \pm 2.0 (0.0; 0.0–12.0)		2.3 \pm 4.6 (0.0; 0.0–30.0)	
>1 CNV (n = 19)	0.2 \pm 0.6 (0.0; 0.0–2.0)	0.022	1.0 \pm 3.3 (0.0; 0.0–13.4)	0.032
≤ 1 LOSS (n = 88)	0.9 \pm 1.9 (0.0; 0.0–12.0)		2.1 \pm 4.4 (0.0; 0.0–30.0)	
>1 LOSS (n = 8)	0.2 \pm 0.6 (0.0; 0.0–1.8)	0.230	1.7 \pm 4.7 (0.0; 0.0–13.4)	0.309
≤ 1 GAIN (n = 92)	0.8 \pm 1.9 (0.0; 0.0–12.0)		2.2 \pm 4.5 (0.0; 0.0–30.0)	
>1 GAIN (n = 4)	0.0 \pm 0.0 (0.0; 0.0–0.01)	0.293	0.0 \pm 0.0 (0.0; 0.0–0.01)	0.29

Sperm concentration and total sperm number are expressed as: mean \pm standard deviation (median; range). Significance is depicted by a p value <0.05.
doi:10.1371/journal.pone.0044887.t006

Table 7. Case-control study of selected losses preliminarily identified by array-CGH as patient-specific (not found in normozoospermic controls).

CNV code	X chr. band	Patients (frequency)	Controls (frequency)	p value	Carriers code	Carriers phenotype (total sperm count $\times 10^6$ and/or testis histology)	Genes inside and nearby (<500 Kb)
17	Xp22.31	2/359 (0.55%)	0/370 (0%)	0.244	A448, A828	A448: Oligozoosperm. (13.4); A828: Oligozoosperm. (20)	NLGN4X, VCX3A ⁵ , HDHD1A [†] , STS
18	Xp22.31	1/359 (0.27%)	0/370 (0%)	0.492	07-96	Azoosperm. (0.0; SCOS)	VCX3A ⁵ , HDHD1A [†] , STS
22	Xp22.11	1/359 (0.27%)	0/370 (0%)	0.492	MMP718	Oligozoosperm. (6.4)	DDX53 ⁵ , PTCHD1 [†] , PRDX4
23	Xp21.3	1/359 (0.27%)	0/370 (0%)	0.492	A142	Cryptozoosperm. (0.01)	POLA1, SCARN23, ARX ⁸
31	Xp11.23	2/270 (0.74%)	0/325 (0%)	0.206	A630, 09-126	A630: Oligozoosperm. (7.2) 09-126: Oligozoosperm. (8)	ARAF, SYN1, TIMP1, CFP, ELK1, UXT, LOC100133957, CXXC1P1, ZNF81, ZNF182 ⁵ , SPACA5 ⁸ , ZNF630, SPACA5B ⁸ , LOC100509575 ⁵ , SSX5 [†] , SSX1, SSX3, SSX4, SSX4B, SLC38A5, FTS11, PORCN, EBP, TBC1D25, RBM3, WDR13
32	Xp11.22	2/359 (0.55%)	0/370 (0%)	0.244	A162, 08-190	A162: Cryptozoosperm. (0.01); 08-190: Azoosperm. (0.0)	MAGED1, SNORA11D, SNORA11E, MAGED4B/MAGED4 ⁵ , XAGE2 [†] , XAGE2B, XAGE1B, XAGE1A, XAGE1D, XAGE1C, XAGE1E
50	Xq22.1	1/359 (0.27%)	0/370 (0%)	0.492	07-22	Azoosperm. (SCOS)	NXF2B, NXF2, TMSB15A, NXF4, ARMCK5, GPRASP1 ⁵ , ARMCK5-GPRASP2 , GPRASP2 [†] , BHLHB9, RAB40A, BEX1, NXF3
54	Xq24	1/359 (0.27%)	0/370 (0%)	0.492	MM550	Oligozoosperm. (0.24)	LONRF3, KIAA1210, PGRMC1 [†] , SLC25A43 [†] , LOC100303728, SLC25A5, CXorf56, UBE2A, NKRIF, SEPT6, MIR766
56	Xq25	1/359 (0.27%)	0/370 (0%)	0.492	06-188	Azoosperm. (0.0; SCOS)	LOC100129520, DCAF12L2 ⁵
57	Xq25	1/359 (0.27%)	0/370 (0%)	0.492	05-238	Cryptozoosperm. (0.22)	DCAF12L2 [†] , DCA12L1
61	Xq27.2	1/359 (0.27%)	0/370 (0%)	0.492	07-30	Azoosperm. (0.0; SCOS)	SPANXA2, SPANXA1, SPANXD, SPANXE ⁵ , MAGEC3 , MAGEC1 [†] , MAGEC2
66	Xq27.3	1/359 (0.27%)	1/370 (0.2%)	1.000	07-516, CS67	07-516: Azoosperm. (0.0; mixed SCOS-hypospermatogenesis); CS67: Normozoosperm. (235)	CXorf1, MIR890, MIR888, MIR892A, MIR892B, MIR891B, MIR891A ⁵
67	Xq28	4/359 (1.11%)	0/370 (0%)	0.058	05-196, MMP676, MMP687, MMP704	05-196, MMP676, 05-196: Azoosperm. (0.0; SCOS); MMP687: Oligozoosperm. (21.5); MMP687: Oligozoosperm. (57.2); MMP704: Oligozoosperm. (1.02).	IDS, LOC100131434, CXorf40A ⁵ , MAGEA9B ⁵ , HSFY2 [†] , HSFY1 [†] , TMEM185A, MAGEA11, HSFY1, HSFY2, MAGEA9, MAGEA8, CXorf40B

Genes inside the CNV minimum size are depicted in bold;

⁵genes inside the CNV maximum size;[†]the first proximal flanking gene;[‡]the first distal flanking gene; the remaining genes are situated <500 Kb from the minimum size border. Azoosperm = Azoospermia; Oligozoosperm = Oligozoospermia; Cryptozoosperm = Cryptozoospermia; SCOS = Sertoli Cell Only Syndrome; SGA = Spermatogenic Arrest.

doi:10.1371/journal.pone.0044887.t007

member E (*SPANXE*). Four deletions (CNV22, 54, 56 and 57) contained several (from 4–32) conserved transcription factor binding sites, but the neighbouring genes were relatively distant (from 8 Kb to 400 Kb).

Discussion

The diffusion of assisted reproductive techniques as a therapeutic option in severe male factor infertility raised several questions about the short and long-term consequences on the offspring, since infertile men are at higher risk of being carriers of genetic anomalies in both their genomic DNA and gametes. Although the importance of diagnosing genetic factors in this category of future fathers is fully recognized, the diagnostic workup of infertile men is still limited to a few genetic tests. Our working hypothesis was that, similarly to Y chromosome-linked CNVs (AZF and gr/gr deletions), we would be able to identify recurrent, pathogenic deletions on the X chromosome. First, an X-chromosome specific high resolution array-CGH analysis was carried out in 199 men with known sperm count and was followed by a screening of selected CNVs in several hundred infertile patients and normozoospermic controls. Our array-CGH analysis showed that 50% of subjects presented at least one CNV, and the majority of these CNVs (55/73) were not reported in currently available databases of genomic variants. Among the few X-linked CNVs reported in subjects with known sperm count [7] only six partially or completely overlapping CNVs were found. This can be due to both technical issues (different array resolution, different criteria used for the interpretation of data, lack of validation in the Tuttelmann paper) and/or due to the patient selection criteria (azoospermic men were selected for a specific histology, called SCOS, in the Tuttelmann et al paper [7]). Interestingly, a small deletion, CNV 69 on Xq28 was observed in 7 patients and 3 controls and it maps inside a CNV reported by Tuttelmann et al [7] as patient-specific, present in a single oligozoospermic German man (“private”). This discrepancy is likely due to the larger size (34 Kb) of the patient-specific deletion in the German patient compared to our 10 subjects (11.7 kb). On the contrary, a reciprocal deletion/duplication (CNV31/CNV30) was observed exclusively in patients (n = 4) in our study, whereas Tuttelmann et al. found two normozoospermic carriers of the duplication and one carrying the deletion [7]. However, the deletion encountered in the above German study was 25 Kb smaller than CNV31/30. An other interesting finding concerns two partially overlapping gains detected in both studies, which affect the dosage of two genes (*H2BFWT* and *H2BFM*). In our study this CNV (CN51) has been found both in controls (n = 4) and patients (n = 5), whereas in the German study [7] it was found only in an oligozoospermic patient. Given that the larger CNV reported in the German study [7] duplicates also two other genes (*TMSB15B*, *H2BFXP*), the combined analysis of the results suggests that it is more likely that the not shared genes, situated in the larger duplication, are responsible for the observed oligozoospermic phenotype.

The further analysis of patient-specific deletions (n = 13) revealed that >90% of them are unique or rare (frequency <1%). These data are in line with the previous whole genome array-CGH study [7] in which among the 27 patient-specific CNVs only one recurrent duplication was found in two oligozoospermic men. Similarly in the paper by Stouffs et al, among the 10 patient specific autosomal CNVs only two were recurrent [8]. The role of rare CNVs has already been established for other multifactorial diseases [18,19] and since mutations causing spermatogenic failure are unlikely transmitted to the next generation, we can predict that *de novo* mutations probably play a

major role in primary testicular failure. It remains difficult to ascertain the importance of rare patient-specific CNVs in spermatogenesis through family analysis, since analysis on maternal X-chromosome would not be informative and brothers (with a 50% chance of sharing the same X chromosome) were not available for analysis. The difficulty to obtain DNA from relatives in relationship with infertility studies is related to the delicate nature of this condition and for this reason the two previous array-CGH studies were also unable to define the *de novo* nature of the identified CNVs. As an alternative way to explore their potential clinical relevance, we performed a search for functional genomic regions (protein coding genes, microRNAs, conserved transcription binding sites) mapping inside or nearby the 13 deletions of interest. Since men are hemizygous for X-linked genes, their CNV-dependent altered expression cannot be compensated by a normal allele and could potentially lead to a direct pathological effect. Ours is the first study suggesting that X-linked CTA family members are recurrently affected and their dosage variation may play a role in CNV-related spermatogenic failure. CTA genes comprise more than 240 members from 70 families and are generally divided into two broad categories: X-linked (mostly multicopy genes) and non-X CTA genes (mainly single copy genes located on autosomes) [for review see [16,20]]. These genes are normally expressed only in germ cells but aberrant activation has also been reported in a number of malignant tumors. The exclusive physiological expression in germ cells strongly suggests a role in spermatogenesis hence human CTA gene family members are largely unexplored and no clinical data is available. Interestingly, by tracing the evolutionary history of CTA genes, it has been demonstrated that CTA genes in general and the X chromosome linked CTA genes in particular are under strong diversifying pressure and amongst the fastest-evolving genes in the human genome [21]. Consequently, many of the human X-linked CTA genes do not have easily identifiable orthologues in the mouse or rat genomes, which makes it difficult to study the role of these genes in animal models. Clues regarding functionality of CTAs for many of these proteins point to a role in cell cycle regulation or transcriptional control [for review see [22]]. Data obtained in the 103 controls (array-CGH analysis) indicates that in this group only one control-specific deletion contained a CTA gene, the sarcoma antigen 1 *SAGE1*, which indicates that this gene is unlikely a spermatogenesis candidate gene. In support of such a statement, the expression of this gene is extremely low in the testis. On the contrary, for the patient-related CTA genes expression levels in the testis and germ cells were substantially higher. Apart from CTA family members we identified other potential candidate genes in the patient group which deserve further genetic screening. On the contrary, we can conclude that those genes which are deleted in control subjects, are unlikely to be spermatogenesis candidate genes since their absence is compatible with normal spermatogenesis. Among the 6 gene-containing control-specific losses, with the exception of vesicle-associated membrane protein 7 (*VAMP7*), the level of testicular expression is either absent or very low. *VAMP7* is situated in PAR2 and it has been described as strongly expressed in the testis, especially in spermatids. Our data indicates that *VAMP7* haploinsufficiency (i.e. one copy of the gene is still retained on the Y-linked PAR2) does not impair spermatogenesis.

One of the most stimulating findings of our article is related to the CNV burden observed in the patients' group in relationship with loss of genetic material. The relatively high frequency of Y chromosome deletions (4–7% in severe spermatogenic failure) already suggested that infertile men are more prone to the loss of genetic material [11]. The mechanism by which Y chromosome

deletions lead to spermatogenetic failure is not fully clarified and they may act either by removing genes involved in spermatogenesis or by affecting meiosis. Here we found an excess of X-linked CNV number and DNA loss in patients with reduced sperm count, which was only partially related to direct gene removal, hence the majority of deletions mapped close to gene-rich areas. We also found a significant association between CNV number and sperm count in the infertile group, which further reinforces the potential link between deletion burden and spermatogenic failure. Similarly to our data, in the paper by Tuttelmann et al [7] a significant inverse correlation has been found between sperm count and CNV number at the whole genome level.

Whether the observed deletions are directly responsible for the phenotype (either affecting gene expression or interfering with sex chromosome pairing for those mapping to the PAR regions) or simply arise due to increased genomic instability, remains a puzzling question. Some previous observations suggest a possible relationship between genomic instability and male infertility and are related to microsatellite instability [23] as well as to the presence of multiple CNVs on the Y chromosome in men with AZF deletions [24] and an excessive CNV number in azoospermic men with SCOS [7]. Previously, we also observed a significant effect of multiple rearrangements in the AZFc region on sperm production, suggesting a potential link between a less stable genome and spermatogenic efficiency [25]. Additionally, epidemiological observations showing a higher incidence of morbidity (including cancer) and lower life expectancy [22,26] in infertile men would support a potential link between altered spermatogenic function and genomic instability. Our study suggests a potential involvement of increased X-linked deletion burden in the aetiology of impaired spermatogenesis and stimulates further research to better define its implication in primary testicular failure and on general health issues for both the patient and his future offspring.

In conclusion, by the analysis of the X chromosome, at the highest resolution available to date, in a large group of subjects with known sperm count we were able to provide evidence about the lack of highly recurrent deletions, which suggest that an AZFc-like region does not exist on this sex chromosome. Our investigation gives an important contribution both to the field of genetics and reproductive medicine since we identified a large number of novel CNVs, and by our second step analysis, we confirmed 12 deletions as being specific to men with impaired spermatogenesis. The analysis of gene-containing CNVs in patients and in controls allows to discern between those that merit future research and those which are unlikely to be involved in spermatogenesis.

Supporting Information

Figure S1 Array-CGH profiles of two CNVs detected by customized oligonucleotide-based X microarray. Magnified

view of CNV 30 (left) and CNV 50 (right) in cases 08-79 and 07-22, respectively. The shaded areas indicate a gain in DNA copy number (duplication, average log2 ratios: +1) detected by red dots (left) and a deletion (average log2 ratios: -4) detected by green dots (right). Arrows indicate the first and the last oligonucleotide duplicated (left) or deleted (right), respectively.

(TIF)

Table S1 List of primers used for the validation of array-CGH results and for the case-control study.

(DOC)

Table S2 List of TaqMan Copy number assay codes used for the validation process.

(DOC)

Table S3 A general outline of the array-CGH findings with phenotypic description of patients and controls.

(XLS)

Table S4 Phenotypic features according to the presence/absence of losses in patients, including the comparison between carriers and no-CNV carriers of hormonal parameters and testis volumes (A) as well as the description of patients with losses detected during both the array-CGH and case-control studies (B).

(DOC)

Table S5 Array-CGH study: comparison of semen parameters according to the number of CNVs in the control group.

(DOC)

Table S6 Array-CGH study: Spermatogenic characteristics of patients and controls carrying more than one CNV.

(DOC)

Acknowledgments

We thank Prof O. Zuffardi, D. Conrad, K. Aston and D. Carrell for helpful discussions. A special thank to Mrs Esperancia Marti from the Fundacio Puigvert for her continues support. We also thank all the clinicians (M. Maggi, A. Magini, F. Lotti) who provided samples for this study from the Andrology Unit of the University of Florence and from the Fundacio Puigvert (L. Bassas, O. Rajmil, J. Sarquella, A. Vives, J. Sanchez-Curbelo).

Author Contributions

Conceived and designed the experiments: CK ER. Performed the experiments: CG DLG FD CC. Analyzed the data: CK ER EA. Contributed reagents/materials/analysis tools: EA ER-C GF. Wrote the paper: CK. Patient recruitment: ER-C GF CK.

References

- Krausz C (2011) Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 25:271–285.
- Nuti F, Krausz C (2008) Gene polymorphisms/mutations relevant to abnormal spermatogenesis. *Reprod Biomed Online* 16:504–513.
- Tuttelmann F, Rajpert-De Meyts E, Nieschlag E, Simoni M (2007) Gene polymorphisms and male infertility—a meta-analysis and literature review. *Reprod Biomed Online* 15:643–658.
- Matzuk MM, Lamb DJ (2008) The biology of infertility: research advances and clinical challenges. *Nat Med* 14:1197–1213.
- Aston KI, Carrell DT (2009) Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *J Androl* 30:711–725.
- Aston KI, Krausz C, Lafaye I, Ruiz-Castane E, Carrell DT (2010) Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. *Hum Reprod* 25:1383–1397.
- Tuttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, et al. (2011) Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PLoS One* 6:e19426.
- Stouffs K, Vandermaelen D, Massart A, Menten B, Vergult S, et al. Array comparative genomic hybridization in male infertility. *Hum Reprod* 27:921–929.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.
- Wang PJ, McCarrey JR, Yang F, Page DC (2001) An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 27:422–426.

11. Krausz C, Chianese C, Giachini C, Guarducci E, Laface I, et al. (2011) The Y chromosome-linked copy number variations and male fertility. *J Endocrinol Invest* 34:376–382.
12. Tyler-Smith C, Krausz C (2009) The will-o'-the-wisp of genetics—hunting for the azoospermia factor gene. *N Engl J Med* 360:925–927.
13. Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, et al. (2008) The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet* 40:794–799.
14. Stouffs K, Tournaye H, Liebaers I, Lissens W (2009) Male infertility and the involvement of the X chromosome. *Hum Reprod Update* 15:623–637.
15. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. *Nature* 444:444–454.
16. Fratta E, Coral S, Covre A, Parisi G, Colizzi F, et al. (2011) The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Mol Oncol* 5:164–182.
17. Del Castillo EB, Trabucco A, De La Balze FA (1947) syndrome produced by absence of the germinal epithelium without impairment of the Sertoli or Leydig cells. *J Clin Endocrinol Metab* 7: 493–502.
18. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753.
19. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466:368–372.
20. Almeida LG, Sakabe NJ, deOliveira AR, Silva MC, Mundstein AS, et al. (2009) CTDdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res* 37:D816–819.
21. Stevenson BJ, Iseli C, Panji S, Zahn-Zabal M, Hide W, et al. (2007) Rapid evolution of cancer/testis genes on the X chromosome. *BMC Genomics* 8:129.
22. Salonia A, Matloob R, Gallina A, Abdollah F, Sacca A, et al. (2009) Are infertile men less healthy than fertile men? Results of a prospective case-control survey. *Eur Urol* 56:1025–1031.
23. Maduro MR, Casella R, Kim E, Levy N, Niederberger C, et al. (2003) Microsatellite instability and defects in mismatch repair proteins: a new aetiology for Sertoli cell-only syndrome. *Mol Hum Reprod* 9:61–68.
24. Jorgez CJ, Weedn JW, Sahin A, Tannour-Louet M, Han S, et al. (2011) Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. *J Clin Endocrinol Metab* 96:E674–679.
25. Krausz C, Giachini C, Xue Y, O'Bryan MK, Gromoll J, et al. (2009) Phenotypic variation within European carriers of the Y-chromosomal gr/gr deletion is independent of Y-chromosomal background. *J Med Genet* 46:21–31.
26. Jensen TK, Jacobsen R, Christensen K, Nielsen NC, Bostofte E (2009) Good semen quality and life expectancy: a cohort study of 43,277 men. *Am J Epidemiol* 170:559–565.

4.2. PAPER 2.

Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility

Lo Giacco D and **Chianese C**, Ars E, Ruiz-Castañé E, Forti G, Krausz C. *J Med Genet.* 2014 May;51(5):340-4.

SHORT REPORT

Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility

D Lo Giacco,^{1,2} C Chianese,³ E Ars,^{1,2} E Ruiz-Castañé,² G Forti,³ C Krausz^{2,3}

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2013-101988>).

¹Molecular Biology Laboratory, Fundació Puigvert Universitat Autònoma de Barcelona, Institut de Investigacions Biomèdiques Sant Pau (IIB-Sant Pau), Barcelona, Spain

²Andrology Service, Fundació Puigvert, Universitat Autònoma de Barcelona, Institut de Investigacions Biomèdiques Sant Pau (IIB-Sant Pau), Barcelona, Spain

³Department of Experimental and Clinical Biomedical Sciences, University of Florence, Italy

Correspondence to

Professor Csilla Krausz, Department of Experimental and Clinical Biomedical Sciences, University of Florence, Viale Pieraccini 6, Florence 50139, Italy; csillagabriella.krausz@unifi.it, c.krausz@dfc.unifi.it

DLG and CC contributed equally to this study.

Received 3 August 2013
 Revised 2 December 2013
 Accepted 27 December 2013

ABSTRACT

Background The role of X-linked genes and copy-number variations (CNVs) in male infertility remains poorly explored. Our previous array-CGH analyses showed three recurrent deletions in Xq exclusively (CNV67) and prevalently (CNV64, CNV69) found in patients. Molecular and clinical characterisation of these CNVs was performed in this study.

Methods 627 idiopathic infertile patients and 628 controls were tested for each deletion with PCR+/- . We used PCR+/- to map deletion junctions and long-range PCR and direct sequencing to define breakpoints.

Results CNV64 was found in 5.7% of patients and in 3.1% of controls ($p=0.013$; OR=1.89; 95% CI 1.1 to 3.3) and CNV69 was found in 3.5% of patients and 1.6% of controls ($p=0.023$; OR=2.204; 95% CI 1.05 to 4.62). For CNV69 we identified two breakpoints, types A and B, with the latter being significantly more frequent in patients than controls ($p=0.011$; OR=9.19; 95% CI 1.16 to 72.8). CNV67 was detected exclusively in patients (1.1%) and was maternally transmitted. The semen phenotype of one carrier (11-041) versus his normozoospermic non-carrier brother strongly indicates a pathogenic effect of the deletion on spermatogenesis. *MAGEA9*, an ampliconic gene reported as independently acquired on the human X chromosome with exclusive physiological expression in the testis, is likely to be involved in CNV67.

Conclusions We provide the first evidence for X chromosome-linked recurrent deletions associated with spermatogenic impairment. CNV67, specific to spermatogenic anomaly and with a frequency of 1.1% in oligo/azoospermic men, resembles the AZF regions on the Y chromosome with potential clinical implications.

more interestingly, we observed a deletion burden in relation to spermatogenic impairment as men with idiopathic infertility had an excessive rate of deletions compared with normozoospermic controls. Among the 29 deletions identified by array-CGH, three recurrent deletions (frequency >1%) on Xq were of interest for their exclusive (CNV67) or prevalent (CNV64 and CNV69) presence in patients. Our previous publication also included a case-control follow-up study in which only 'patient-specific' CNVs were incorporated, and CNV67 was found in 4/359 (1.1%) patients and 0/370 controls.

To evaluate the potential role of these recurrent X-linked CNVs in male infertility, we screened more than 1200 men with known sperm parameters in two Mediterranean populations. All three deletions have been confirmed as significant risk factors for impaired spermatogenesis. In particular, CNV67 was confirmed as 'patient-specific', being the first X-linked deletion with potential clinical implications.

METHODS

Subjects

Germline DNA from 1255 subjects (627 strictly selected patients with idiopathic infertility and 628 normozoospermic controls) from Spain (36.6%) and Italy (63.4%) was analysed (for patient selection criteria see also online supplementary materials). The ethnic/geographical composition was similar between the control and patient groups. The clinical features of the patients are presented in online supplementary table S1.

Molecular analysis of deletions

A multistep Sequence Tagged Site (STS) PCR+/- protocol was optimised in order to identify and confirm reliably the presence of deletions (see online supplementary table S2). PCR+/- was also used to further restrict the deletion interval for CNV67 and CNV64.

Pedigree analysis

Segregation analysis in relatives was possible for CNV67 carriers 11-041 (Spanish) and MMP704 (Italian). To understand whether CNV67 occurred *de novo*, we analysed each patient's mother and siblings (11-041's brother and MMP704's sister). Screening of the female relative of CNV67 carriers was performed by quantitative PCR (qPCR) using a TaqMan Copy Number Assay (hs03323870_cn).

CNV69 deletion breakpoint definition

Conventional PCR+/- using primers mapping to the flanking regions of the minimum CNV69 size

INTRODUCTION

The aetiology of altered spermatogenesis remains unknown in about 40% of cases (so-called 'idiopathic infertility'), of which a large proportion are probably related to still unknown genetic factors.¹ During recent years, copy-number variations (CNVs) have been shown to be an interesting aspect also in andrology. To date, the only known CNVs that actually cause spermatogenic failure are Y chromosome microdeletions; however, high-resolution whole genome approaches have enabled the identification of new spermatogenesis candidate genes as well as recurrent and private patient-specific CNVs with potential clinical interest also on autosomes and the X chromosome.²⁻⁴ In a previously published study,⁴ based on high-resolution X chromosome-specific array-CGH platforms (average resolution 4 kb), we provided the largest collection of X-linked CNVs related to spermatogenesis and,

To cite: Lo Giacco D, Chianese C, Ars E, *et al.* *J Med Genet* Published Online First: [please include Day Month Year] doi:10.1136/jmedgenet-2013-101988

Copy-number variation

was performed to confine breakpoints to smaller intervals. Long-range (LR) PCR was subsequently performed to amplify the junction fragment including the breakpoint (figure 1A; see online supplementary table S3). Additional primers were then designed to sequence the obtained LR-PCR product (figure 1B). To easily classify the type of breakpoint in all CNV69 carriers and to provide a tool for a potential diagnostic screening, another pair of deletion-specific primers was designed (see online supplementary table S3).

Analysis of sequence family variants (SFVs)

In order to understand whether CNV67 caused the deletion of the *MAGEA9* or *CXorf40A* gene copy mapping within the CNV maximum size, we tested carriers for a number of STSs, mapping within the *MAGEA9* and *CXorf40A* ampliconic regions. Each of these markers amplifies two homologous sequences containing sequence family variants (SFVs) that would allow, prior to sequencing, the distinction between the copies inside and outside the maximum size (see online supplementary tables S4 and S5).

Statistical analysis

Statistical analysis was performed using SPSS software V17.0. Significance was tested using the Fisher exact test and corrected by the Holm test for multiple testing (see online supplementary methods).

RESULTS

Physical mapping and bioinformatic characterisation

All three deletions map to the long arm of the X chromosome in q27.3 (CNV64) and q28 (CNV67 and CNV69).

CNV64

This deletion, described in the Database of Genomic Variants (nsv829407), removes between 3.923 and 6.382 kb of DNA, considering its minimum (chrX: 143 436 346–143 440 268) and maximum (chrX: 143 434 786–143 441 167) CNV size, respectively. The presence of highly repetitive sequences in this region prevented fine mapping of the deletion breakpoints. Nevertheless, based on the \pm STS pattern, we localised proximal and distal breakpoints within a region of 0.6 kb upstream and 0.3 kb downstream of the proximal and distal edges of the minimum CNV size, respectively. No genes and regulatory elements were directly removed by this deletion. Nevertheless, a number of functional elements are located at <0.5 Mb from the deletion and may be affected (table 1).

CNV69

This deletion removes between 11.770 and 22.141 kb of DNA, considering its minimum (chrX: 154 044 876–154 056 645) and maximum (chrX: 154 037 065–154 059 205) CNV size, respectively. Sequencing showed the existence of at least two

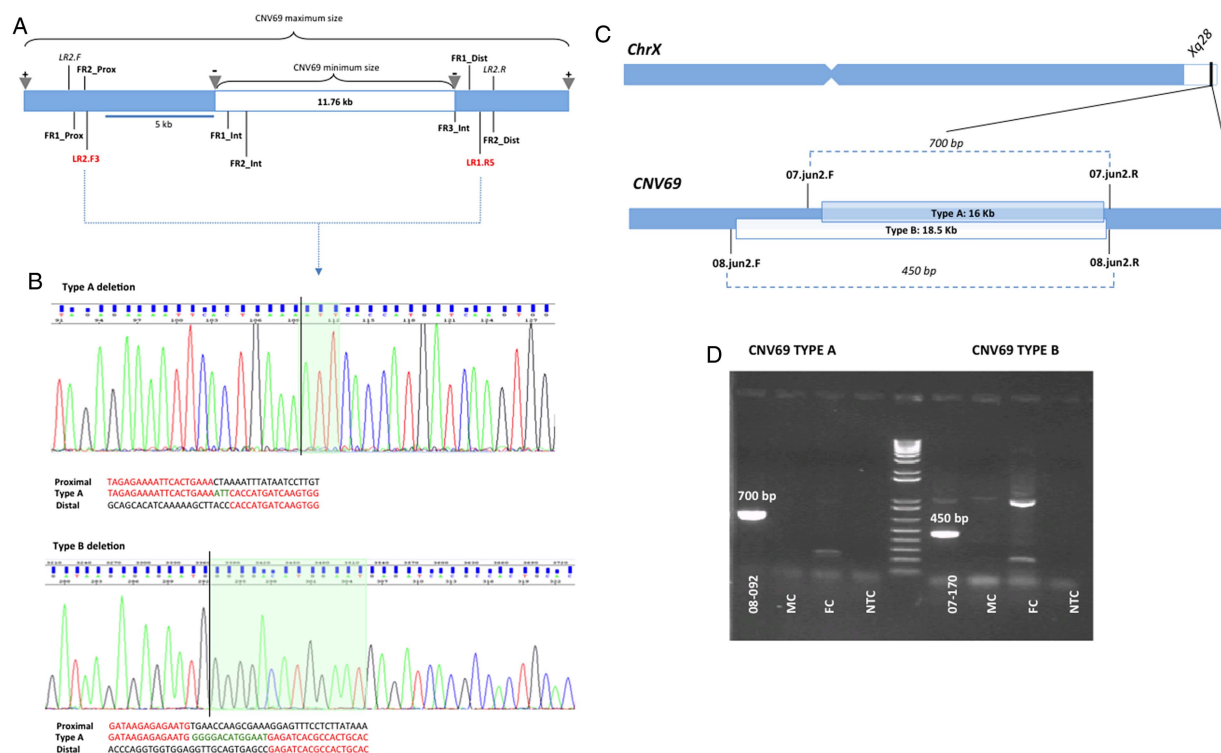


Figure 1 Fine mapping of copy-number variation 69 (CNV69) deletion breakpoint. (A) Schematic representation of the minimum and maximum CNV size. Triangles (+) depict the first and last positive array-CGH probes, delineating the maximum CNV size, and triangles (–) depict the first and last negative array-CGH probes, delineating the minimum CNV size. Primers used for chromosome walking are indicated in bold. Prox, primers located upstream of the proximal edge of the minimum CNV size; Dist, primers located downstream of the distal edge of the minimum CNV size; Int, primers inside the minimum CNV size. Primers used for junction amplification and sequencing are indicated in italic and red, respectively. (B) Sequencing of deletion junction in type A and type B breakpoints. Sequences shown were obtained with LR.R5 reverse primer. Insertions compared with the reference sequence at the deletion junction are highlighted in green and the alignment of the sequences flanking the breakpoint is shown in the lower panel of each electropherogram. (C) Type of deletions. Breakpoint-specific primers used for screening of CNV69 carriers are indicated in bold. Dotted lines depict the amplification product obtained for each type of deletion. (D) PCR product using breakpoint-specific primers. Electrophoresis of the two fragments obtained, each mirroring the deletion sizes. No amplification was obtained in male and female controls. MC, male control; FC, female control; NTC, no template control.

Table 1 Regulatory elements detected within CNV64, CNV67 and CNV69

CNV code	Position	Number of regulatory elements*						Insulator
		Weak enhancer	Strong enhancer	Weak promoter	Active promoter	Weak transcribed region	Transcription elongation	
64	Upstream proximal edge	10	–	–	–	2	–	4
	Inside the deletion	–	–	–	–	–	–	–
	Downstream distal edge	11	3	1	–	2	–	7
67	Upstream proximal edge	57	19	7	3	11	4	8
	Inside the deletion	6	2	–	–	1	–	–
	Downstream distal edge	70	17	10	4	16	1	16
69	Upstream proximal edge	110	38	21	7	28	7	16
	Inside the deletion	–	–	–	–	–	–	–
	Downstream distal edge	68	9	25	7	8	2	4

Only regulatory elements located at <500 kb from the proximal/distal maximum CNV size border are considered.

*Predicted functional elements in normal four human cell lines from ENCODE (cell lines: GM12878-B lymphocyte, lymphoblastoid, International HapMap Project, H1-hESC-embryonic stem cells, HUVEC-umbilical vein endothelial cells, NHER-epidermal keratinocytes).

CNV, copy-number variation.

different deletion patterns referred to as type A and type B (figure 1C). The alignment of the flanking sequences indicates that none of the deletion types originates from homologous recombination, so the most likely mechanism is Non-Homologous End Joining (NHEJ). This hypothesis is further supported by the presence of different additional nucleotides (NHEJ ‘molecular scars’) compared with the reference sequence at the deletion junction (figure 1B). For instance, type A breakpoint was characterised by the insertion of a TAA tract (InsTAA), implying a deletion length of 16.06 kb, whereas type B breakpoint displayed the insertion of a CATTCCATGTCCCC tract (InsCATTCCATGTCCCC) and the deletion measured 18.53 kb. Breakpoint-specific primers allowed the detection of two fragments, each mirroring the deletion sizes: 700 bp for type A and 450 bp for type B (figure 1C,D). Of 32 subjects (22 patients and 10 controls) carrying CNV69, 17 displayed type A (10 patients and 7 controls) and 10 displayed type B (9 patients and 1 control). In the remaining five subjects (3 patients and 2 controls), the deletion could not be classified as either type A or type B and was thus referred to as type C.

No genes are mapping within the deletion. However, the region at <0.5 Mb surrounding the maximum CNV size includes a number of regulatory elements and several genes, of which *BRCC3* was reported as being associated with reproductive phenotypes (tables 1 and 2).⁵

CNV67

Based on array-CGH analysis, we previously reported that this deletion removes between 5.417 and 25.513 kb of DNA, considering its minimum (chrX: 148 456 473–148 461 889) and maximum (chrX: 148 452 726–148 478 238) CNV size based on array-CGH, respectively.⁴ Although the exact deletion breakpoints could not be defined, chromosome walking allowed us to better define the deletion extension, estimated to be 11.664 kb. It is worth noticing that the highly repetitive nature of this region may predispose to inversions,⁶ and thus we propose two possible scenarios (see online supplementary figure S1). Also, in this case the most likely mechanism for deletion formation is NHEJ. The molecular characterisation also highlighted that the proximal maximum size has been underestimated with array-CGH, probably due to the repetitive nature of the region. The distal deletion breakpoint was restricted to a 15.261 kb region downstream of the distal edge of the minimum size, which includes the proximal copy of *MAGEA9* and is 3.7 kb from the proximal copy of *HSPX1/2*. Furthermore, the region at <0.5 Mb from the proximal and distal edges of the maximum CNV size also includes other genes and regulatory elements (tables 1 and 2). The results of the SFV analysis in the *CXorf40A* ampliconic region are shown in online supplementary figure S2. Carrier 10-314 was homozygous for units 1 and 2, the more proximal to the CNV minimum size. Homozygosity was also detected in unit 3 inside the *CXorf40A* gene whereas the SFV

Table 2 Case–control study and list of genes mapping inside and within the flanking regions of CNV64, CNV67 and CNV69

X chromosome band	CNV code	Patients (frequency)	Controls (frequency)	p Value (OR; 95% CI)	Genes inside and nearby
Xq27.3	64	36/627 (5.7%)	19/628 (3.03%)	0.013 (1.89; 1.10 to 3.27)	–
Xq28	67	7/627 (1.1%)	0/628 (0.0%)	0.008	<i>IDS</i> ; <i>LINC00893</i> ; <i>CXorf40A</i> *; <i>MAGEA9B</i> ; <i>HSPX2†</i> ; <i>HSPX1†</i> ; <i>TMEM185A</i> ; <i>MAGEA11</i> ; <i>HSPX1</i> ; <i>HSRX2</i> ; <i>MAGEA9B</i> ; <i>MAGEA</i> ; <i>CXorf40B</i>
Xq28	69	22/627 (.5%)	10/628 (1.59%)	0.033 (2.2; 1.05 to 4.62)	<i>GAB3</i> ; <i>DKC1</i> ; <i>SNORA36A</i> ; <i>SNORA56</i> ; <i>MPP1</i> ; <i>SMIM9</i> ; <i>F8</i> ; <i>H2AFB3FUND2</i> ; <i>CMC4</i> ; <i>MTCP1</i> ; <i>BRCC3*</i> ; <i>VBP1†</i> ; <i>RAB39B</i> ; <i>CLIC2</i> ; <i>TMLHE-AS1</i> ; <i>H2AFB3</i> ; <i>F8A1</i> ; <i>F8A1</i> ; <i>H2AFB1</i> ; <i>TMLHE-AS1</i>
	Type A	10/615 (1.6%)	7/624 (1.1%)	0.301 (NS)	
	Type B	9/614 (1.5%)	1/619 (0.2%)	0.011 (9.19; 1.16 to 72.8)	
	Type C	3/608 (0.5%)	2/620 (0.3%)	0.491 (NS)	

Genes potentially removed by the deletion are highlighted in bold. Only genes located <500 kb from the proximal/distal maximum CNV size border are considered.

*First proximal flanking gene.

†First distal flanking gene.

NS, not significant.

Copy-number variation

pattern changed to heterozygosity for the outer units (ie, units 4/5 and 6). Interestingly, of the five SFVs mapping within this unit, 10-314 never displayed the variant mapping inside the CNV maximum size. The rest of the carriers were heterozygous for different SFVs and, accordingly, different types of deletion have been defined (see online supplementary figure S2). As for the analysis of the *MAGEA9* ampliconic region, carriers were mostly homozygous for the analysed SFVs; however, the majority of SFVs were uninformative due to the fact that homozygosity was also observed in female and male controls. However, the comparison of *MAGEA9* SFVs between carrier 11-041 and his non-carrier brother showed that the carrier was homozygous for all SFVs analysed whereas his brother was heterozygous for all SFVs except one (see online supplementary table S6).

Genotype–phenotype correlation

All three deletions were significantly more frequent in patients than in controls (table 2). The phenotype of the carriers is shown in online supplementary table S7. Estimating CNV69 deletion frequencies according to the type of breakpoint, we observed that type B breakpoint was significantly more frequent in patients (9/614, 1.5%) than in controls (1/619, 0.2%) ($p=0.011$; OR=9.19, 95% CI 1.16 to 72.8). No statistically significant difference was found between patients (10/615, 1.6%) and controls (7/624, 1.1%) (table 2) with regard to type A breakpoint.

CNV67 was found in 1.1% of patients and in no controls ($p=0.008$ (significant $p<0.017$ after correcting by the Holm test; table 2).

All CNV carriers displayed a significantly lower total sperm count and total motile sperm count compared with non-carriers, considering the whole study population (see online supplementary table S8A), but not when comparison was performed in the two phenotypic groups separately (see online supplementary table S8B). A comparison of the clinical features of carriers and non-carriers is presented in online supplementary table S9.

Pedigree analysis

CNV67 was maternally inherited, and neither of the patients' mothers nor MMP704's sister (who also carried CNV67) had premature ovarian failure or anovulation. Interestingly, patient 11-041's brother did not carry CNV67 and was normozoospermic, different from his oligozoospermic brother who did carry CNV67 (see online supplementary figure S3).

DISCUSSION

In this study we provide evidence for an association of CNV64, CNV67 and CNV69 with spermatogenic failure through (1) a case–control study on a large study population; (2) molecular characterisation of deletions; (3) a search for functional elements in the region of interest; and (4) genotype–phenotype correlation analysis.

Concerning the patient-enriched CNVs CNV64 and CNV69, two clues support their association with impaired spermatogenesis: (1) deletion carriers had an increased probability of impaired spermatogenesis compared with non-carriers (OR=1.9 and 2.2, respectively); and (2) semen quality in terms of total sperm count and total motile sperm count was significantly impaired in carriers compared with non-carriers. Type B CNV69 (the larger one) was significantly more represented in patients than in controls, suggesting that this deletion pattern may account for the potential deleterious effect of CNV69 on sperm production. This may be related to the closer position of type B deletion (7.8 kb) to an upstream insulator compared with

type A, the proximal breakpoint of which maps to 2.6 kb downstream (see online supplementary figure S4). Importantly, through the breakpoint definition we developed a simple diagnostic tool for type B deletion, allowing other laboratories to further explore the role of this deletion in other ethnic groups.

We confirmed in more than 1200 subjects the recurrent (1.1%) and 'patient-specific' feature of CNV67. Unfortunately, it was impossible to obtain a fine mapping of CNV67 breakpoints because of the presence of highly repetitive sequences and the incomplete assembly of the currently available reference sequence of the human X chromosome derived from 16 different individuals.⁶ SFV analysis provided a more accurate estimate of the minimum and maximum deletion size of CNV67—for instance, when heterozygosity is observed, deletion is surely not present. Specifically, SFV analysis in the *CXorf40A* ampliconic region suggests that different deletion breakpoints might exist at the proximal edge of the CNV minimum size (see online supplementary figure S2), determining the involvement of the *CXorf40A* copy mapping inside the maximum size. Accordingly, one of our patients (carrier 10-314) displays the extent of the larger deletion that would remove part of the *CXorf40A* coding sequence. With regard to SFV analysis in the *MAGEA9* ampliconic region, the variants are not fully informative. However, the homozygous pattern observed for most variants in our carriers allows speculation that the proximal copy of the *MAGEA9* gene might be involved. It is especially evident in the case of the two brothers, given that the deletion carrier showed homozygosity whereas his brother was heterozygous for the SFVs mapping to the *MAGEA9* copy. Although we were unable to formally demonstrate the removal of the proximal copy of the *MAGEA9* gene on the basis of the above results, we can speculate that this gene is directly affected or the deletion affects its regulatory elements. For instance, large-scale CNVs might change the three-dimensional structure of chromatin, which is seemingly crucial for correct gene regulation.^{7–9}

We previously suggested that Cancer Testis gene dosage variation may play a role in CNV-related spermatogenic failure⁴; accordingly, *MAGEA9* belongs to this gene family. *MAGEA9* is an ampliconic gene reported as independently acquired on the human X chromosome, since no orthologs could be detected in the murine X chromosome.⁶ Independently acquired X-linked genes are predominantly expressed in the testis with a specific expression of multi-copy genes in male germ cells.⁶ It can therefore be speculated that the loss of *MAGEA9* copies would affect spermatogenesis. Moreover, according to gene ontology, *MAGEA9* is mainly involved in the regulation of gene expression, DNA methylation, reproduction and spermatogenesis, reflecting its potential involvement in transcriptional and epigenetic regulatory mechanisms of gametogenesis.⁵ Finally, CNV67 may also affect the regulation of *HSFY1/2*, another independently acquired X-linked multi-copy gene with testis-specific expression.

Pedigree analyses of two CNV67 carriers indicated that this deletion is maternally inherited, thus not affecting female fertility. This is in accordance with the lack of expression of *MAGEA9* in the ovary. The family of patient 11-041 is especially informative since the pathological semen phenotype of the carrier (11-041) versus his normozoospermic non-carrier brother is a strong indicator for a pathogenic effect of the deletion on spermatogenesis.

For the first time we provide evidence for a significant association between recurrent X-linked deletions and impaired sperm production. Strikingly, CNV67, which is specific to spermatogenic failure, resembles AZF deletions on the Y chromosome.

This finding merits further investigations in order to elucidate the structural complexity of this region and to provide a feasible substrate for fine molecular characterisation and large-scale diagnostic testing.

Acknowledgements The authors wish to thank the clinicians from the Andrology Unit of the University of Florence (M Maggi, A Magini, F Lotti), the Fundació Puigvert (L Bassas, O Rajmil, J Sarquella, A Vives, J Sanchez-Curbelo) and the University of Ancona (G Balercia) who helped in providing samples for this study. We also thank Dr H Skaletsky and Dr F Daguin for useful discussions and E Rossi (University of Pavia, Italy) for her input on array-CGH data. Special thanks to Esperança Martí, President of the Fundació Puigvert, for her continuous support. The IIB Sant Pau-Fundació Puigvert Biobank is acknowledged for kindly providing part of the DNA samples.

Contributors All authors are justifiably credited with authorship, according to the authorship criteria.

Funding This work was funded by grants from the Spanish Health Ministry FIS 11/02254 and Italian Ministry University (Grant PRIN 2010-2012 to CK).

Competing interests None

Patient consent Obtained.

Ethics approval The local ethical committees of the University Hospital Careggi and the Fundació Puigvert approved the study.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 Krausz C. Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 2011;25:271–85.
- 2 Tüttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, Wieacker P, Röpke A. Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PLoS ONE* 2011;6:e19426.
- 3 Lopes AM, Aston KI, Thompson E, Carvalho F, Gonçalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A, Seabra C, Wilfert AB, Dai J, Downie JM, Fernandes S, Guo X, Sha J, Amorim A, Barros A, Carracedo A, Hu Z, Hurler ME, Moskvovtsev S, Ober C, Paduch DA, Schiffman JD, Schlegel PN, Sousa M, Carrell DT, Conrad DF. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS Genetics* 2013;9:e1003349.
- 4 Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, Ars E, Ruiz-Castañe E, Forti G, Rossi E. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PLoS ONE* 2012;7:e44887.
- 5 Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, Demougin P, Gattiker A, Schulze W, Jégou B, Kirchhoff C, Primig M. Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum Reprod* 2012;27:3233–48.
- 6 Mueller JL, Skaletsky H, Brown LG, Zaghul S, Rock S, Graves T, Auger K, Warren WC, Wilson RK, Page DC. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* 2013;45:1083–7.
- 7 Sanyal A, Lajoie BR, Jain G, Decker J. The long-range interaction landscape of gene promoters. *Nature* 2012;489:109–13.
- 8 Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, Rozowsky J, Birney E, Bickel P, Snyder M, Gerstein M. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol* 2012;13:R48.
- 9 Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addelman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Fietze S, Fu Y, Gertz J, Grubert F, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012;488:91–100.

4.3. PAPER 3.

X chromosome-linked CNVs in male infertility: discovery of overall duplication load and recurrent, patient-specific gains with potential clinical relevance.

Chianese C, Gunning AC, Giachini C, Daguin F, Balercia G, Ars E, Lo Giacco D, Ruiz-Castañé E, Forti G, Krausz C. *PLoS One*. 2014 Jun 10;9(6):e97746.



X Chromosome-Linked CNVs in Male Infertility: Discovery of Overall Duplication Load and Recurrent, Patient-Specific Gains with Potential Clinical Relevance

Chiara Chianese^{1,3}, Adam C. Gunning¹, Claudia Giachini¹, Fabrice Daguin¹, Giancarlo Balercia², Elisabet Ars³, Deborah Lo Giacco^{3,4}, Eduard Ruiz-Castañé⁴, Gianni Forti¹, Csilla Krausz^{1,4*}

1 Department of Experimental and Clinical Biomedical Sciences, University of Florence and Centre of Excellence DeNothe, Florence, Italy, **2** Division of Endocrinology, Department of Clinical and Molecular Sciences, Umberto I Hospital, Polytechnic University of Marche, Ancona, Italy, **3** Molecular Biology Laboratory, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain, **4** Andrology Service, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona, Spain

Abstract

Introduction: Spermatogenesis is a highly complex process involving several thousand genes, only a minority of which have been studied in infertile men. In a previous study, we identified a number of Copy Number Variants (CNVs) by high-resolution array-Comparative Genomic Hybridization (a-CGH) analysis of the X chromosome, including 16 patient-specific X chromosome-linked gains. Of these, five gains (DUP1A, DUP5, DUP20, DUP26 and DUP40) were selected for further analysis to evaluate their clinical significance.

Materials and Methods: The copy number state of the five selected loci was analyzed by quantitative-PCR on a total of 276 idiopathic infertile patients and 327 controls in a conventional case-control setting (199 subjects belonged to the previous a-CGH study). For one interesting locus (intersecting DUP1A) additional 338 subjects were analyzed.

Results and Discussion: All gains were confirmed as patient-specific and the difference in duplication load between patients and controls is significant ($p = 1.65 \times 10^{-4}$). Two of the CNVs are private variants, whereas 3 are found recurrently in patients and none of the controls. These CNVs include, or are in close proximity to, genes with testis-specific expression. DUP1A, mapping to the PAR1, is found at the highest frequency (1.4%) that was significantly different from controls (0%) ($p = 0.047$ after Bonferroni correction). Two mechanisms are proposed by which DUP1A may cause spermatogenic failure: i) by affecting the correct regulation of a gene with potential role in spermatogenesis; ii) by disturbing recombination between PAR1 regions during meiosis. This study allowed the identification of novel spermatogenesis candidate genes linked to the 5 CNVs and the discovery of the first recurrent, X-linked gain with potential clinical relevance.

Citation: Chianese C, Gunning AC, Giachini C, Daguin F, Balercia G, et al. (2014) X Chromosome-Linked CNVs in Male Infertility: Discovery of Overall Duplication Load and Recurrent, Patient-Specific Gains with Potential Clinical Relevance. PLoS ONE 9(6): e97746. doi:10.1371/journal.pone.0097746

Editor: Joël R. Drevet, Clermont-Ferrand Univ., France

Received: December 6, 2013; **Accepted:** April 24, 2014; **Published:** June 10, 2014

Copyright: © 2014 Chianese et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Tuscan Regional Health Research Program; Spanish Ministry of Health (grant FIS-11/02254); PRIN 2009; European Commission - Reproductive Biology Early Research Training (REPROTRAIN, Project Number: 289880). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: c.krausz@dfc.unifi.it

Introduction

Infertility is a multi-factorial disorder affecting approximately 15% of couples – half of these can be attributed to the male. Currently known causes of male-factor infertility account for only 60% of cases and known genetic factors contribute to about 15% of severe male factor infertility [1]. The most frequent molecular genetic cause is related to the Y chromosome and concerns the AZF deletions [2]. These deletions are the first example in andrology of functionally-relevant CNVs and can be easily studied with plus/minus PCR. Recently, the development of high-throughput analytical techniques such as a-CGH have allowed the screening of large numbers of loci and have been used with the principal aim of identifying novel spermatogenesis candidate genes. These studies have also been useful in identifying a CNV burden in infertile men, mainly involving the sex chromosomes [3–5].

Considering the high complexity of spermatogenesis, which requires more than 2,000 genes, it is highly likely that a proportion of the 40% ‘missing’ aetiology is linked to yet unknown genetic factors [1]. CNVs may induce a pathogenic effect in a number of ways: structural changes to regulatory regions or a numerical increase or decrease in protein-coding regions may have a direct effect on mRNA levels [6]; large-scale CNVs may cause changes to the well-regulated 3D structure formed by chromatin [7], leading to downstream effects on the regulation of protein-coding regions. Finally, large CNVs may also disturb chromosome pairing at the PAR regions during meiosis [8,9].

While the AZF region-linked genes have been extensively studied in respect to male infertility [10] very few studies have focussed on the X chromosome, despite its predicted enrichment in genes expressed in the testis [11,12]. Only a single X-linked gene has been shown to definitively contribute to an infertility

phenotype, the androgen receptor (*AR*) [1], leaving huge scope for further experiments on this chromosome. Recently, evolutionary models demonstrated that the X chromosome progressively accumulating testis-expressed genes [11]. Accordingly, 18% of the 800 protein-coding genes on the human X chromosome do not show orthologs in other placental mammals. Of these, the majority have been acquired by the human X chromosome *de novo* or through transposition from the autosomes. Two-thirds of these are ampliconic, possessing duplicated 10 Kb segments with >99% homology and are predicted to be involved in male fitness [12].

In our previous study [3], we analyzed the CNV status of 96 infertile patients and 103 controls using a custom-designed 8×60 K microarray targeting the X chromosome (Agilent Technologies, Santa Clara, CA, USA). Of the 44 gains identified, 16 were patient-specific and the five most promising CNVs (DUP1A, DUP5, DUP20, DUP26 and DUP40) were selected for testing in an enlarged study population.

Materials and Methods

The local Ethics Committees of the University Hospital Careggi and the Fundació Puigvert approved the study and consent procedure. All participants gave written, informed consent. The consent forms are stored locally at the University Hospital Careggi. All data were analysed anonymously.

CNV Selection and Bioinformatic Analysis

Five CNVs (DUP1A, DUP5, DUP20, DUP26 and DUP40) were selected from the 16 patient-specific gains identified in our previous study [3]. CNVs underwent several selection steps. Initially, this focused on the frequency at which CNVs were identified in the a-CGH study. CNVs found in control samples were excluded. Online data-sources, such as OMIM, gene ontology terms, and literature review were used to find candidate features within 0.5 Mb of the CNV minimum. Information about expression data was obtained from microarray and RNA-seq experiments deposited in the GermOnline database [13,14]. CNVs containing genomic features with a potential involvement in spermatogenesis were selected. At this regard, DUP1A resulted of particular interest and therefore it was subjected to a deepened investigation. All genomic *loci*, in the previous study reported in Hg18 genome assembly [3], have been converted to Hg19 assembly and all *loci* in the current study are reported according to Hg19.

Study Population

Patients with non-Italian or non-Spanish origin were excluded. Patients were recruited after a comprehensive clinical examination in both centres using the same clinical protocol, according to WHO guidelines [15]. Exclusion criteria included all known causes of infertility: history of mono- or bilateral cryptorchidism, varicocele (grades 2 and 3), obstructive azoospermia, recurrent infections, iatrogenic infertility, hypogonadotrophic hypogonadism, karyotype anomaly and Y-chromosome microdeletions. Control samples were recruited from pre-vasectomy patients with proven fertility, male partners of infertile women (tubal factor, anovulation and endometriosis) and male volunteers from the general population. Given that the aim of the study was to define the effect of CNVs on spermatogenesis, the control group contained only subjects presenting with normal sperm parameters (according to the WHO manual). Additional information about sperm parameters is provided in the supplementary material and methods S1 in File S1.

In addition to the previous study population of 199 samples (96 patients and 103 controls) [3], 404 new subjects (180 patients and 224 controls) were analyzed at the 5 selected loci. The total patient group ($n = 276$) used in this study consisted of 96 azoospermic, 83 cryptozoospermic ($<1.0 \times 10^6$ spermatozoa. ml^{-1}) and 97 severe oligozoospermic subjects (between 1.0×10^6 and 5.0×10^6 spermatozoa. ml^{-1}) of which 130 were Spanish and 146 Italian. The control group ($n = 327$) consisted of 107 Spanish and 220 Italian normozoospermic subjects. For DUP1A only, the further enlargement of the study population with additional 158 patients (27 azoospermic, 59 cryptozoospermic and 72 severe oligozoospermic) and 180 controls ensured the analysis of a total of 941 subjects.

qPCR Analysis

The copy number state of each locus was determined by qPCR. Primer sequences are reported in Table S2 in File S1. Before testing, all samples were put through rigorous quality control to ensure that DNA quality and concentration was sufficient. DNA samples that showed contamination were re-precipitated, using ethanol precipitation. DNA samples were diluted to $50 \text{ ng} \cdot \mu\text{L}^{-1}$ in ddH₂O. Each sample was analyzed in triplicate in a 96-well plate. The SYBR Select Master Mix produced by Invitrogen (REF: 4472908) was used. For each sample, *PMP22* was amplified as a reference gene for analysis purposes. The reaction conditions were as follows: 20 ng DNA; 200 nM Primer (forward and reverse); SYBR Green SELECT Master mix (1x concentration) in a total reaction volume of 20 μL . In the case of DUP1A, an additional qPCR was performed in order to test the *LINC00685/PPP2R3B* gene dosage ratio. For this purpose, we tested the *PPP2R3B* copy number state in DUP1A carriers using a pair of primers solely mapping to the *PPP2R3B* gene. We used the same reaction conditions, but optimal concentration for these primers was 400 nM (forward and reverse). qPCR was performed on TaqMan 7900 HT on 'Absolute Quantification', using the pre-set 'Standard' cycle conditions. The annealing temperature was 60°C. A non-targeting control and a 'duplicated' control (a Klinefelter 47, XXY man) was included on each plate. For Dup1A, we simultaneously analysed sample 08–373 (carrying DUP4A) and A800 (carrying DUP1A) for whom CNVs had previously been identified by a-CGH [3]. Threshold cycle and baseline were calculated automatically and relative quantification was determined using the $\Delta\Delta\text{Ct}$ analysis method.

All samples from the original a-CGH experiment (96 patients and 103 controls) were re-tested using the qPCR method validating a-CGH results. A single 'normal' result (absence of duplication in the triplicate) was required to denote a normal gene copy number. In case duplication was found, the analysis was repeated once for confirmation. A 'borderline' range was also possible and samples within this range were repeated; if a borderline result turned into duplication, another experiment was performed for confirmation.

Statistical Analysis

SPSS software (version 20.0, Chicago, IL, USA) was used. CNV frequencies were analysed for significance using Fisher's Exact Test in the 'R' software package and corrected using Bonferroni-Holm step-down correction for multiple testing.

Results

Case-control Association Study

The 5 selected CNVs (DUP1A, DUP5, DUP20, DUP26 and DUP40) were confirmed as patient-specific. Among the 276 patients analyzed, DUP5 was found in two (0.72%), DUP20 and

DUP26 in one (0.36%), DUP40 in 3 (1.09%) and DUP1A in 4 patients (1.44%). Given the higher frequency of DUP1A, additional samples were tested for this CNV. In this enlarged study population, we found DUP1A in a further two patients (2/158; 1.26%) and in none of the additional 180 controls. Considering the total study population used for DUP1A only (434 patients and 507 controls), the difference in duplication frequency between patients and controls for this CNV reached statistical significance after Bonferroni-Holms correction for multiple testing (6/434 patients versus 0/507 controls; $p = 0.047$) (Table 1). DUP1A carriers displayed a heterogeneous semen phenotype ranging from azoospermia to severe oligozoospermia (Table 2). A phenotypic description of carriers of all CNVs is shown in Table 2.

A comparison of sperm parameters between all CNV carriers and non-carriers was performed and is shown in Table S1 in File S1. The most relevant results concern the Total Motile Sperm count, which resulted significantly lower in carriers of DUP1A ($p = 0.008$), DUP26 ($p = 0.003$) and DUP40 ($p = 0.03$) compared to non-carriers.

Bioinformatic Analysis for Physical Characterization of CNVs

The physical characteristics of the selected CNVs are shown in Table 3. DUP1A and DUP5 were of special interest because of their large size and location on the PAR1. Locations of DUP20, DUP26 and DUP40 are Xp22.2, Xp21.1 and Xq21.1, respectively. No homology of sufficient size for non-allelic homologous recombination (NAHR) was found between the upper and lower boundaries of all CNVs indicating that a mechanism other than NAHR might have led to the formation of these recurrent gains. We checked for other types of repeated elements that could contribute to genome instability, such as *Alu* elements, and found that, for DUP1A, DUP5, DUP20 and DUP26, the flanking regions are filled with short interspersed nuclear elements (SINEs), which include *Alus*, and long interspersed nuclear elements (LINEs). As for DUP40, two LINE elements, belonging to the L1PA3 family1, are located at the extremities of the CNV. Hence, we propose that the presence of SINEs and LINEs might underlie the generation of recurrent duplications.

Gene Content of Interest and Search in the Database of Genomic Variants (DGV)

DUP1A fully duplicates the following genes: *PLCXD1*, *GTPB6* and *LINC00685* (Table 3). Apparently, the most interesting is a long non-coding RNA, (*LINC00685*) predicted to act as a negative regulator of a gene (*PPP2R3B*) with a potential role in spermatogenesis (Figure 1). According to the a-CGH data, the *PPP2R3B* gene is affected by DUP1A only for an 11.7 Kb span,

thus it is not duplicated but disrupted at intron 7–8 by the CNV (considering the minimum size). Similarly, qPCR data confirm the lack of complete duplication of this gene in all CNV carriers.

Through DGV search within DUP1A, several CNVs (30 duplications and 25 deletions referring to the variant esv27600) were found intersecting the *PLCXD1* and *GTPBP6* loci (located in the proximal part of the duplication), and all of them were identified only in females without further notification on the phenotype. Four variants are reported in close proximity to the most interesting genes, *LINC00685* and *PPP2R3B*. As for *LINC00685*, two distal variants are reported in DGV (4 Kb downstream), esv2219721 and esv266687: the former refers to a 549 bp deletion found in only one man, whereas the latter refers to a 560 bp loss found in 82/1151 (7.1%) subjects (including men and women) analyzed. As for the *PPP2R3B* gene, three variants are reported in DGV and none of them affects the entire gene. The first (esv27923) refers to a 548 bp loss mapping to intron 10–11 of the gene found in one woman over a total of 451 subjects (0.22%); the second (esv25834) refers to a 2.12 Kb region mapping to intron 7–8 of the gene, where both gains ($n = 7$) and losses ($n = 4$) were found in 11 women over a total of 451 subjects (2.24%); the third (esv2758560) refers to 3 gains (218.7 Kb) mapping to intron 1–2 of the gene, found in three men over 271 subjects totally analysed (1.1%). All these CNVs seemingly do not disturb the ratio between *LINC00685* and *PPP2R3B* copy number, as with DUP1A.

Within the CNV minimum of DUP5 there is a single predicted microRNA element, *AL732314.1* (Table 3). For DUP5, the already described esv2758560 variant is reported in DGV as well as a large number of small CNVs; one variant (nsv508745), describing three insertions (one in a woman and the other two in one man out of 270 belonging to the HapMap project), overlaps with the *AL732314.1*. No information about fertility of the single male carrier could be found.

Within the CNV minimum of DUP20, two protein-coding genes are duplicated entirely (Table 3). Of interest is *MSL3*, an homologue of the *Drosophila melanogaster* homonymous gene, which is involved in X-chromosome dosage compensation [16]. A number of CNV annotations can be found in DGV. Of the 3 merged variants (nsv523223: 2.4 Kb; nsv524154: 74.9 Kb; nsv526302: 140.8 Kb) representing 3 gains spanning the *MSL3* locus, no CNVs were found in males. The ten (nsv515163, esv2739963, esv270995, nsv6797, esv271693, nsv510814, nsv6799, esv2739964, nsv436626, nsv499329) common CNVs found upstream of *MSL3* are all found at a much higher frequency in females than in males (28 women (1.66%) versus 10 men (0.59%) over a total of 1,682 subjects reported in the DGV).

Both DUP26 and DUP40 do not contain any protein-coding genes but of interest is the presence of *FAM47C*, located 200 Kb downstream of DUP26, and *HMGV5*, located 139 Kb upstream of

Table 1. CNV frequency and statistical analysis of case-control association study.

CNV	Frequency Patients	Frequency Controls	Raw p-value	Corrected p-value ¹
DUP1A	6/434 (1.38%)	0/507	0.01	0.047
DUP5	2/276 (0.72%)	0/327	0.21	0.63
DUP20	1/276 (0.36%)	0/327	0.45	0.45
DUP26	1/276 (0.36%)	0/327	0.45	0.45
DUP40	3/276 (1.09%)	0/327	0.097	0.39

¹Corrected using Bonferroni-Holm Step-down correction for multiple testing.
doi:10.1371/journal.pone.0097746.t001

Table 2. Phenotypic Description of CNV carriers.

CNV	Carrier Code	Sperm Conc. (10^6 Spzoa. mL^{-1})	Total Sperm Count (10^6 Spzoa)	Total Motile Sperm Count (10^6 Spzoa)	FSH (U.L^{-1})	LH (U.L^{-1})	Testosterone (nmol.L^{-1})	Mean testicular volume ³ (mL)
DUP1A	05-002	1.40	2.1	0.48	11.6	5.34	24.2	11
	08-093	0.18	0.4	0.1	9.63	N ¹	N/A ²	14
	08-280	1.83	8.5	0.9	10.2	3.71	21	15
	11-262	0.0	0.0	0.0	26.3	11.5	20.8	14
	13-099	4.00	16.0	1.28	5.94	3.58	17.5	18
	A800	0.50	2.0	0.36	4.91	3.8	16.2	14
DUP5	08-280	1.83	8.5	0.9	10.2	3.71	21	15
	A760	0.0	0.0	0.0	3.23	4.64	7.1	15
DUP20	M9	5.00	7.5	3.6	4.5	3.5	N ¹	12
DUP26	07-013	0.0	0.0	0.0	10	2.2	N/A ²	18
DUP40	05-238	0.22	0.22	0.0	5.0	N ¹	N/A ²	13.5
	07-002	0.0	0.0	0.0	4.1	2.8	15.6	15
	A238	0.01	0.01	0.0	6.5	4.3	N ¹	14

¹Reported in the medical history as within the "normal range".²Not Available, but according to medical history "No signs of hypoandrogenism".³Testis volume was determined using the 'Prader' orchidometer;

Spzoa = Spermatozoa.

doi:10.1371/journal.pone.0097746.t002

Table 3. Physical Characteristics of CNVs selected for the study.

CNV	Start-end position (CNV Min) ¹	Size (Min)	Size (Max)	Substrate for NAHR ²	Protein Coding within, or nearby CNV minimum (within 0.5 Mb)	Regulatory/RNA within, or nearby CNV minimum (within 0.5 Mb)
DUP1A	ChrX: 61544–306372 ¹	245 Kb	247 Kb	No	<i>PLCXD1</i> , <i>GTPBP6</i> , <i>PPP2R3B</i> ⁴ , <i>SHOX</i>	<i>LINC00685</i> , <i>AL732314.1</i>
DUP5	ChrX: 382644–542740 ¹	160 Kb	168 Kb	No ³	<i>PLCXD1</i> , <i>GTPBP6</i> , <i>PPP2R3B</i> , <i>SHOX</i>	<i>LINC00685</i> , <i>AL732314.1</i> , <i>RP11–309M23.1</i>
DUP20	ChrX: 11194597–11796693 ¹	602 Kb	608 Kb	No ³	<i>MID1</i> , <i>HCCS</i> , <i>ARHGAP6</i> , <i>AMELX</i> , <i>MSL3</i> , <i>FRMPD4</i> , <i>PRPS2</i> , <i>TLR7</i> , <i>TLR8</i> , <i>TMSB4X</i> , <i>FAM9C</i> .	<i>RP11–120D5.1</i> , <i>AC002981.1</i> , <i>FRMPD4-AS1</i> , <i>TLR8-AS1</i> , <i>RP11–791M20.1</i> , <i>GS1–600G8.5</i>
DUP26	ChrX: 37283466–37372045 ¹	89 Kb	96 Kb	No ³	<i>FAM47C</i> , <i>PRRG1</i> ⁴ , <i>TM4SF2</i> ⁴ , <i>LANCL3</i> , <i>XX</i> , <i>CYBB</i> , <i>DYNLT3</i> , <i>CXorf27</i> , <i>SYTL5</i> .	<i>RNU6–49</i>
DUP40	ChrX: 80225590–80230870 ¹	5.3 Kb	28 Kb	No	<i>BRWD3</i> , <i>HMG5</i> , <i>SH3BGR</i>	<i>ACA64</i> , <i>U6</i> , <i>AL357115.1</i>

¹Genomic positions are in Hg19.²Non-allelic Homologous Recombination (NAHR).³Short regions (<300 bp) with 80% homology were found, but were not considered sufficient for NAHR.⁴Gene crossed the minimum and maximum threshold, and is not fully duplicated.

doi:10.1371/journal.pone.0097746.t003

DUP40 (Table 3). Both these genes show testis-specific expression [14]. Concerning DUP26, three variants are reported to overlap with this CNV: esv1007820 (6.9 Kb), describing a gain found in a man without information on semen parameters; esv28598 (6.3 Kb), describing a gain found in a woman over a total of 451 subjects; esv2740093 (13.6 Kb), describing a loss found in one man over 96 subjects totally analyzed. None of these CNVs intersect the *FAM47C* directly but are located between 307.7–325.6 Kb downstream. Two losses (esv23127, size: 380.5 Kb; nsv510839, size: 74.4 Kb) and 1 gain (esv24190, size: 18.1 Kb) covered the entire DUP40 CNV minimum, whereas one gain (nsv436916, size: 10.26 Kb) covered DUP40 for most of its size (3.8 Kb). Again, these CNVs do not intersect *HMG5* directly but are located between 238.3–281.1 Kb upstream. The 2 gains and the 2 losses were found at the same frequency in 2 women and 2 men over a total of 508 analyzed subjects (0.39%). No information on the men's fertility status was available.

Discussion

A recent study by Tüttelmann et al [4] provides the first statistically significant duplication burden on the X chromosome, reporting a significantly higher number of gains in azoospermic patients compared to both oligozoospermic patients and normal controls. Our analysis of five X-linked CNVs also reveals a

significantly higher duplication load in infertile compared to normozoospermic men.

While four of the five CNVs (DUP5, DUP20, DUP26 and DUP40) studied did not individually reach statistical significance, they remained patient-specific. It is worth noting that rare variants have previously been predicted to play an important role in spermatogenic failure [5] due to the strong selection against highly-penetrant infertility-causing variants. All CNVs include, or are in close proximity to, genes with testis-specific expression and potential implication in spermatogenesis. DUP20 contains the *MSL3* gene, a homologue of *Drosophila melanogaster* homonymous gene. *Msl3* in *D. melanogaster* plays a critical role in the X-chromosome dosage-compensation pathway by directing histone H4 acetylation at lysine 16 (H4K16) and the human homologue is thought to have a similar function [17]. Although no information in humans is available, mice models provide evidence that this specific chromatin modification is dramatically increased in elongating spermatids and precedes histone replacement during spermatogenesis, as an initial step of nucleosome removal [18]. Both DUP26 and DUP40 are within close proximity of genes expressed exclusively in the testis. Inside and nearby DUP40 there is a dense area with epigenetic features indicating that DUP40 may disrupt the epigenetic regulation of neighbouring genes. DUP5 is of particular interest due to its large size and location on the PAR1 as explained below.

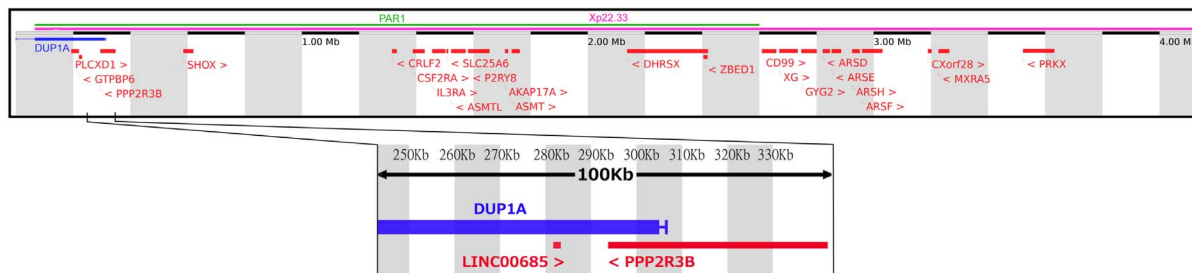


Figure 1. Position of DUP1A relative to *LINC00685* and *PPP2R3B* on the X chromosome (PAR1). Diagram of the Xp22.33, showing the presence of all known protein-coding genes (Red) and DUP1A (Blue). Enlarged is a 100 Kb region showing the location of *PPP2R3B* and *LINC00685* in relation to the DUP1A minimum and maximum. This gain will certainly duplicate the antisense element *LINC00685*, but does not fully duplicate *PPP2R3B* – skewing the ratio between the gene and its negative regulator.

doi:10.1371/journal.pone.0097746.g001

DUP1A was found at a significantly higher frequency in patients. This gain contains a long non-coding RNA (*LINC00685*) that potentially acts as a negative regulator of a gene with potential role in spermatogenesis, *PPP2R3B*. This proposed mechanism could not be confirmed by functional studies because *in vitro* human spermatogenic cell culture is not available and this antisense is not present in easily accessible model organisms such as mouse and *Drosophila*. However, expression data provides evidence for an inverse relationship between *PPP2R3B* and *LINC00685* levels in a number of different tissues (see Figure S1 in File S1 and Figure 2) [13,14]. Concerning the testis, those samples with a high expression of *PPP2R3B* show comparatively low expression of *LINC00685* [13,14].

Figure 2 shows the changing levels of *PPP2R3B* and *LINC00685* mRNA throughout spermatogenesis and the inverse relationship between *PPP2R3B* and *LINC00685*. In tissues that lack germ cells, *PPP2R3B* levels are low, and *LINC00685* levels are comparatively high (AdMinus, JS1 and JS2). *PPP2R3B* levels rise with the presence of mitotically active cells (AdPlus, JS3), and are highest in samples enriched in meiotically active cells (JS5). Scoring system for testicular biopsies was obtained from Chalmel et al [14] and is reported in Table S3 in File S1. Again, in these tissues *LINC00685* levels are decreased. Based on this observation, we propose that the mechanism by which DUP1A could lead to spermatogenic failure is through increased negative regulation, caused by the duplicated *LINC00685* that would decrease *PPP2R3B* transcription in the developing germ cells. This hypothesis is also supported by our qPCR analysis, which proved that the six patients with DUP1A, thus carrying an over-dosage of the antisense *LINC00685*, do not display a duplication of the entire *PPP2R3B* gene.

Although the role of *PPP2R3B* in spermatogenesis has not been explored, indirect functional evidence supports its involvement both in mitosis and meiosis. *PPP2R3B* encodes for a subunit of the protein phosphatase 2 (PP2) protein complex. PP2 is one of four major Ser/Thr phosphatases and has been implicated in a wide range of cellular activities. PP2 is a heterotrimeric protein, composed of a structural A-subunit, a catalytic C-subunit and a regulatory B-subunit. *PPP2R3B* belongs to this final group, giving specificity to the PP2 complex [19]. In mice, *PPP2R3B*-PP2 was demonstrated to maintain a pool of dephosphorylated CDC6 - a replication-licensing factor [19]. CDC6 phosphorylation and dephosphorylation is necessary for the mitotic G₁ to S-phase transition [20]. Accordingly, overexpression of *PPP2R3B* results in mitotic arrest at G₁ phase [19]. Unlike somatic cells, following mitosis primary spermatocytes retain high levels of CDC6 [20], indicating a possible further role for this protein in meiosis.

Expression analysis of testis biopsies with different histology patterns supports the involvement of *PPP2R3B* in mitosis where *PPP2R3B* levels are higher in samples containing mitotically active cells compared to those lacking germ cells. More importantly, the most pronounced expression is observed in biopsies with spermatocytic arrest i.e. those enriched in meiotic cells, indicating an additional role for this gene in meiosis [14]. As mentioned above, *LINC00685* levels show an opposite trend, strongly suggesting a functional link between the two gene products [13,14] (see supplementary discussion S2, Table S3 and Figure S1 in File S1).

Given that genes that have been recently incorporated into the human X chromosome are involved in male reproductive fitness [12] it is worth noting that *PPP2R3B* has been recently acquired on this chromosome.

Although no CNVs were found in the database of genomic variants (DGV) intersecting *PPP2R3B* or *LINC00685*, re-analysis of raw data deposited in dbVar by Tüttelmann et al. [4] shows that a number of CNVs (5 gains and 2 losses) were found in the PAR1 of azoospermic men. Focusing on the gains, the variant nsv869733 (243.8 Kb), describing 3 gains mapping to the *PPP2R3B* locus, was found exclusively in azoospermic patients (n = 4). The CNV minimums of two of these gains (one mapping to chrX: 298, 292–322, 672 and the other mapping to chrX: 298, 292–330, 801) begin in *PPP2R3B* intron 1–2 and end in intron 12–13, respectively. These CNVs are likely to lead to a non-functional protein by causing an internal duplication. The other CNV (mapping to chrX: 291, 285–336, 040) begins 3.4 Kb upstream of the *PPP2R3B* locus and ends in intron 12–13, which would disrupt the gene at exon 1 as well as the promoter. Moreover, the presence of a predicted promoter 1Kb upstream of *PPP2R3B* and several regions enriched in histone methylation suggests that these gains may disrupt the correct regulation of *PPP2R3B*. The duplication of the *PPP2R3B* locus has been reported also in men with abnormal karyotype and AZF deletions and in a single man with normal karyotype and AZFb deletion [21]. However, in these cases the infertile phenotype is clearly related to the karyotype and Y chromosome defects. The observations from Tüttelmann, together with ours, showing the relatively high frequency at which DUP1A was found exclusively in patients and its potential link to *PPP2R3B* gene expression, strongly indicate that *PPP2R3B* could be considered a novel spermatogenesis candidate gene.

Large CNVs of the PAR1 region, like DUP1A and DUP5, may lead to impaired spermatogenesis also through a structural effect disturbing male meiosis i.e. altering the recombination event that occurs between the PAR1 regions of the sex chromosomes [8,9]. PAR1 recombination becomes progressively more frequent

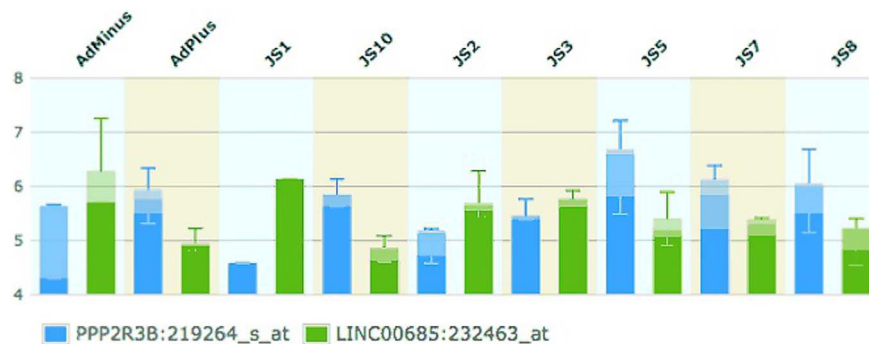


Figure 2. Relative expression of *PPP2R3B* and *LINC00685* in testes biopsies in patients with different phenotypes (Johnsen Score) (Chalmel, et al., 2012). Taken from EBI Expression Atlas. (<http://www.ebi.ac.uk/gxa/home>). doi:10.1371/journal.pone.0097746.g002

towards the distal telomeric boundary [22], where DUP1A and DUP5 are located, showing the importance of this region during meiosis. For instance, the merged analysis of data by Tüttelmann and our previous study shows that no PAR-linked CNVs were found in the vicinity of these 2 CNVs in normozoospermic controls, with the exception of one subject found by Krausz et al. to be carrying DUP4A. However, this CNV only partially overlaps with the DUP1A (at its distal boundary) and DUP5 (at its proximal boundary). This observation may indicate that: i) the pathogenic effect of DUP1A is more likely related to the misbalanced *LINC00685/PPP2R3B* gene dosage effect; ii) the portions of DUP1A and DUP5 that are not overlapping with DUP4A could be affecting specific sites of importance for X-Y pairing during meiosis.

Finally, the combination of all data discussed above supports the importance of the PAR1-linked CNVs in male infertility. Our most relevant finding is the identification of the first recurrent, X-linked gain associated with spermatogenic failure, DUP1A. Two possible mechanisms have been provided to explain the pathogenesis of the associated infertile phenotype – one identifying and regarding a novel spermatogenesis candidate gene and another

due to a potential structural effect in the PAR1. Both of these scenarios are intriguing and prompt further research.

Supporting Information

File S1 Materials and Methods: S1. **Table S1.** Impact of CNVs on Total Sperm Count (TSC) and Total Motile Sperm Count (MSC) Comparison was performed between carriers and non-carriers, excluding controls (A) and including controls (B). **Table S2.** Primers used for qPCR Analysis. **Discussion:** S2. **Table S3** - Scoring system for testicular biopsies (Johnsen score). **Figure S1.** Relative expression of PPP2R3B and LINC00685 in Human organs taken from EBI Expression Atlas. (DOC)

Author Contributions

Conceived and designed the experiments: CK CC AG. Performed the experiments: CC AG CG DLG. Analyzed the data: CC AG CK FD. Contributed reagents/materials/analysis tools: GF CK EA ERC GB. Wrote the paper: CC AG CK.

References

- Krausz C (2011) Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 25: 271–285. doi:10.1016/j.beem.2010.08.006.
- Krausz C, Hoefsloot L, Simoni M, Tüttelmann F (2014) EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: state-of-the-art 2013. *Andrology* 2: 5–19. doi:10.1111/j.2047-2927.2013.00173.x.
- Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, et al. (2012) High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PLoS One* 7: e44887. doi:10.1371/journal.pone.0044887.
- Tüttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, et al. (2011) Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PLoS One* 6: e19426. doi:10.1371/journal.pone.0019426.
- Lopes AM, Aston KI, Thompson E, Carvalho F, Gonçalves J, et al. (2013) Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and Both Sex Chromosomes, including the Gene DMRT1. *PLoS Genet* 9: e1003349. doi:10.1371/journal.pgen.1003349.
- Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 10: 451–481. doi:10.1146/annurev.genom.9.081307.164217.
- Sanyal A, Lajoie BR, Jain G, Dekker J (2012) The long-range interaction landscape of gene promoters. *Nature* 489: 109–113. doi:10.1038/nature11279.
- Chandley AC, Edmond P (1971) Meiotic studies on a subfertile patient with a ring Y chromosome. *Cytogenetics* 10: 295–304.
- Mohandas TK, Speed RM, Passage MB, Yen PH, Chandley AC, et al. (1992) Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. *Am J Hum Genet* 51: 526–533.
- Krausz C, Chianese C, Giachini C, Guarducci E, Laface I, et al. (2011) The Y chromosome-linked copy number variations and male fertility. *J Endocrinol Invest* 34: 376–382.
- Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, et al. (2008) The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet* 40: 794–799. doi:10.1038/ng.126.
- Mueller JL, Skaletsky H, Brown LG, Zaghul S, Rock S, et al. (2013) Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* 45: 1083–1087. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23872635>. Accessed 12 December 2013.
- Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SSW, Demougin P, et al. (2007) The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A* 104: 8346–8351. doi:10.1073/pnas.0701883104.
- Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, et al. (2012) Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum Reprod* 27: 3233–3248. doi:10.1093/humrep/des301.
- WHO (2010) World Health Organization Laboratory Manual for the Examination and Processing of Human Semen.
- Prakash SK, Van den Veyver IB, Franco B, Volta M, Ballabio A, et al. (1999) Characterization of a novel chromo domain gene in xp22.3 with homology to *Drosophila msl-3*. *Genomics* 59: 77–84. doi:10.1006/geno.1999.5844.
- Smith ER, Cayrou C, Huang R, Lane WS, Côté J, et al. (2005) A human protein complex homologous to the *Drosophila* MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol* 25: 9175–9188. doi:10.1128/MCB.25.21.9175-9188.2005.
- Lu L-Y, Wu J, Ye L, Gavrilina GB, Saunders TL, et al. (2010) RNF8-dependent histone modifications regulate nucleosome removal during spermatogenesis. *Dev Cell* 18: 371–384. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2840054&tool=pmcentrez&rendertype=abstract>. Accessed 24 February 2014.
- Yan Z, Fedorov SA, Mumby MC, Williams RS (2000) PR48, a novel regulatory subunit of protein phosphatase 2A, interacts with Cdc6 and modulates DNA replication in human cells. *Mol Cell Biol* 20: 1021–1029.
- Eward KL, Obermann EC, Shreeram S, Loddo M, Fanshawe T, et al. (2004) DNA replication licensing in somatic and germ cells. *J Cell Sci* 117: 5875–5886. doi:10.1242/jcs.01503.
- Jorgez CJ, Weedon JW, Sahin A, Tannour-Louet M, Han S, et al. (2011) Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. *J Clin Endocrinol Metab* 96: E674–9. doi:10.1210/jc.2010-2018.
- Filatov DA (2004) A gradient of silent substitution rate in the human pseudoautosomal region. *Mol Biol Evol* 21: 410–417. doi:10.1093/molbev/msh032.

4.4. PAPER 4.

Y-chromosome microdeletions are not associated with SHOX haploinsufficiency.

Chianese C, Lo Giacco D, Tüttelmann F, Ferlin A, Ntostis P, Vinci S, Balercia G, Ars E, Ruiz-Castañé E, Giglio S, Forti G, Kliesch S, Krausz C. *Hum Reprod.* 2013 Nov;28(11):3155-60.

Y-chromosome microdeletions are not associated with SHOX haploinsufficiency

C. Chianese¹, D. Lo Giacco^{2,3}, F. Tüttelmann⁴, A. Ferlin⁵, P. Ntostis¹, S. Vinci¹, G. Balercia⁶, E. Ars², E. Ruiz-Castañé^{2,3}, S. Giglio⁷, G. Forti⁸, S. Kliesch⁹, and C. Krausz^{1,3,*}

¹Andrology Unit, Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence 50139, Italy, ²Molecular Biology Laboratory, Fundació Puigvert, Universitat Autònoma de Barcelona, Barcelona 08025, Catalonia, Spain, ³Andrology Service, Fundació Puigvert, Barcelona 08025, Catalonia, Spain, ⁴Institute of Human Genetics, University of Münster, Münster 48149, Germany, ⁵Department of Molecular Medicine, Section of Clinical Pathology and Centre for Human Reproduction Pathology, University of Padova, Padova 35128, Italy, ⁶Division of Endocrinology, Department of Clinical and Molecular Sciences, Umberto I Hospital, Polytechnic University of Marche, Ancona 60121, Italy, ⁷Medical Genetics Unit, Meyer Children's University Hospital, Florence 50139, Italy, ⁸Endocrinology Unit, Department of Experimental and Clinical Biomedical Sciences, University of Florence, Florence 50139, Italy, and ⁹Centre of Reproductive Medicine and Andrology, University Clinics, Münster 48149, Germany

*Correspondence address. E-mail: c.krausz@dfc.unifi.it

Submitted on May 24, 2013; resubmitted on July 5, 2013; accepted on July 16, 2013

STUDY QUESTION: Are Y-chromosome microdeletions associated with *SHOX* haploinsufficiency, thus representing a risk of skeletal anomalies for the carriers and their male descendants?

SUMMARY ANSWER: The present study shows that *SHOX* haploinsufficiency is unlikely to be associated with Y-chromosome microdeletions.

WHAT IS KNOWN ALREADY: Y-chromosome microdeletions are not commonly known as a major molecular genetic cause of any pathological condition except spermatogenic failure. However, it has been recently proposed that they are associated not only with infertility but also with anomalies in the pseudoautosomal regions (PAR), among which *SHOX* haploinsufficiency stands out with a frequency of 5.4% in microdeletion carriers bearing a normal karyotype. This finding implies that sons fathered by men with Y-chromosome defects will not only exhibit fertility problems, but might also suffer from *SHOX*-related conditions.

STUDY DESIGN: Five European laboratories (Florence, Münster, Barcelona, Padova and Ancona), routinely performing Y-chromosome microdeletion screening, were enrolled in a multicenter study.

PARTICIPANTS/MATERIALS, SETTING, METHODS: PAR-linked and *SHOX* copy number variations (CNVs) were analyzed in 224 patients carrying Y-chromosome microdeletions and 112 controls with an intact Y chromosome, using customized X-chromosome-specific array-CGH platforms and/or qPCR assays for *SHOX* and *SRY* genes.

MAIN RESULTS AND THE ROLE OF CHANCE: Our data show that 220 out of 224 (98.2%) microdeletion carriers had a normal *SHOX* copy number, as did all the controls. No *SHOX* deletions were found in any of the examined subjects (patients as well as controls), thus excluding an association with *SHOX* haploinsufficiency. *SHOX* duplications were detected in 1.78% of patients ($n = 4$), of whom two had an abnormal and two a normal karyotype. This might suggest that Y-chromosome microdeletions have a higher incidence for *SHOX* duplications, irrespective of the patient's karyotype. However, the only clinical condition observed in our four *SHOX*-duplicated patients was infertility.

LIMITATIONS, REASONS FOR CAUTION: The number of controls analyzed is rather low to assess whether the *SHOX* duplications found in the two men with Y-chromosome microdeletions and a normal karyotype represent a neutral polymorphism or are actually associated with the presence of the microdeletion.

WIDER IMPLICATIONS OF THE FINDINGS: Men suffering from infertility due to the presence of Y-chromosome microdeletions can resort to artificial reproductive technology (ART) to father their biological children. However, infertile couples must be aware of the risks implied and this makes genetic counseling a crucial step in the patient's management. This study does not confirm previous alarming data that showed an association between Y-chromosome microdeletions and *SHOX* haploinsufficiency. Our results imply that deletion carriers have no augmented

risk of *SHOX*-related pathologies (short stature and skeletal anomalies) and indicate that there is no need for radical changes in genetic counseling of Yq microdeletion carriers attempting ART, since the only risk established so far for their male offspring remains impaired spermatogenesis.

STUDY FUNDING/COMPETING INTEREST(S): This work was supported by the Italian Ministry of University (grant PRIN 2010-2012 to C.K.), Tuscan Regional Health Research Program ('Progetto Salute 2009') to G.F., the Spanish Ministry of Health (grant FIS-11/02254) and the European Union 'Reprotrain' Marie Curie Network (project number: 289880 to C.K.). The authors declare that no conflicting interests exist.

Key words: Y chromosome / *SHOX* / male infertility / spermatogenesis / gr/gr deletion

Introduction

The Y chromosome displays a distinctive genomic landscape for most of its extent, rich in repetitive elements that provide a favorable environment for the generation of copy number variations (CNVs). There is ever-growing evidence that collocates CNVs among those genetic factors that deserve prime attention when dealing with complex human diseases. To date, the only CNVs unquestionably associated with male infertility are the Y-chromosome microdeletions, which involve the azoospermia factor (AZF) region on Yq and are thus termed AZFa, AZFb and AZFc microdeletions (Krausz et al., 2011). Each type of microdeletion is in a clear-cut cause–effect relationship with a distinct abnormal semen phenotype, but until recently microdeletion carriers with normal karyotype have not been proven to be at risk for any condition other than infertility.

However, one recent study (Jorgez et al., 2011) reported that microdeletion carriers also displayed aberrations in the pseudoautosomal regions (PAR1 and PAR2), short homologous regions sited at the extremities of the gonosomes. Therefore, the authors proposed that the mechanism underlying Y-chromosome microdeletions might also be associated with the occurrence of PAR rearrangements. The salient, and alarming, finding of this study resided in the detection of haploinsufficiency of the PAR1-located *SHOX* (Short stature HOmeobox-containing) gene in 5.4% of men carrying Yq microdeletions and a normal karyotype. It is widely ascertained that *SHOX* haploinsufficiency leads to disproportionate short stature and diverse skeletal anomalies such as Leri–Weill dyschondrosteosis (LWD) (Helena Mangs and Morris, 2007). In contrast, the occurrence of *SHOX* duplications is apparently rare, with only few cases reported so far (Grigeliene et al., 2001; Tachdjian et al., 2008; Roos et al., 2009; Thomas et al., 2009; D'Haene et al., 2010; Gervasini et al., 2010) and no direct relationship with any specific phenotype has yet been defined. The finding by Jorgez et al. (2011) raised the question whether microdeletion carriers might be at higher risk of incurring PAR-related pathologies. If this were true, genetic counseling would radically change, since infertile couples undergoing artificial reproductive technology (ART) would need to be informed that their sons not only will have fertility problems but also will be at risk of developing PAR-related disorders. Moreover, screening for *SHOX*-linked CNVs should then become compulsory for men carrying Yq microdeletions.

Given the relevance and the potential clinical impact of this issue, we performed a multicenter investigation on a large study population—almost doubling that of the above mentioned study (Jorgez et al., 2011)—in order to investigate whether the hypothesis of an association between Yq microdeletions and *SHOX* haploinsufficiency could be confirmed.

Materials and Methods

Subjects

The study population counted a total of 336 Caucasian men, comprising 224 patients carrying different types of Y-chromosome microdeletions (4 complete AZFa; 6 AZFb,c; 153 complete AZFc; 4 AZFa,b,c; 57 partial AZF deletions; 40 gr/gr, 7 gr/gr-b2/b4 duplication, 3 partial AZFa deletion, 5 b2/b3, 1 b1/b3, 1 b3/b4) and 112 men with an intact Y chromosome, referred to as controls. The detection of Y-chromosome microdeletions was achieved by screening patients according to the European Academy of Andrology (EAA) guidelines (Simoni et al., 2004), in the following participating laboratories: Florence ($n = 66$); Münster ($n = 56$); Barcelona ($n = 43$); Padova ($n = 49$); Ancona ($n = 10$). Karyotype was available for 300 subjects, including patients and controls.

Germline DNA deriving from peripheral blood lymphocytes was originally isolated in each participating center by standard methods. DNA quality was assessed using a Nanodrop ND-1000 UV-VIS spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA) and samples showing a A_{260}/A_{280} ratio > 1.8 were used. All individuals gave informed consent for the routine AZF analysis and further scrutiny.

Array-comparative genomic hybridization

Customized X-chromosome-specific array-comparative genomic hybridization (CGH) platforms (8X60 K, Agilent Technologies, Santa Clara, CA, USA) were generated using the e-Array software (<http://earray.chem.agilent.com/>); 53 069 probes (60-mer oligonucleotides) were selected from the Agilent database and covered the whole X chromosome, including Xp and Xq PARs, with a mean resolution of 4 kb. Four replicate probe groups, with each probe present in two copies on the platform, were designed in regions containing mouse infertility-associated genes, i.e. sperm protein associated with the nucleus, X-linked family members (SPANX); testis expressed 11 (*TEX11*), TAF7-like RNA polymerase II (*TAF7L*), TATA box binding protein (*TBP*). In these regions, the medium resolution is 2 kb. For the normalization of copy number changes, the array also included Agilent control clones spread along all autosomes (6842 probes). As a reference DNA, the same male subject with no Yq microdeletions and normal karyotype was used. This control DNA was already characterized for CNV content in previous array-CGH experiments against eight different normozoospermic controls and presented one private gain of 27 kb mapping to Xcentr which was not considered for the frequency analyses. Three hundred nanograms of test DNA and control DNA were double-digested with *RsaI* and *AluI* (Promega, Madison, WI, USA) for 1 h at 37°C. After digestion, samples were incubated at 65°C for 20 min to inactivate the enzymes, and then labeled by random priming (Agilent Technologies, Santa Clara, CA, USA) for 2 h using Cy5-dUTP for the test DNA and Cy3-dUTP (Agilent Technologies, Santa Clara, CA, USA) for the control DNA. Labeled DNAs were incubated at 65°C for 10 min and then purified with Microcon YM-30 filter units (Millipore, Billerica, MA, USA). Every purified sample

was brought to a total volume of 9.5 ml in 1xTE (pH 8.0, Promega, Madison, WI, USA), and yield and specific activity were determined for each sample using a NanoDrop ND-1000 UV-VIS Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). The appropriate cyanine 5- and cyanine 3-labeled samples were combined in a total volume of 16 ml. After sample denaturation and pre-annealing with 5 µl of Human Cot-1 DNA (Invitrogen, Carlsbad, CA, USA), hybridization was performed at 65°C with shaking for 24 h. After two washing steps, the array was analyzed through the Agilent scanner and the Feature Extraction software (v10.1.1). Graphical overview was obtained using the DNA Analytics (v4.0.73). All the array experiments were analyzed using the ADM-2 algorithm at threshold 5. Aberrant signals including three or more adjacent probes were considered as genomic CNVs. The positions of oligomers refer to the Human Genome March 2006 assembly (hg18). Array-CGH calls for the detected CNVs are provided in [Supplementary data, Table S1](#) and raw data are depicted in [Supplementary data, Figs A–D](#), referring to samples presented in Table 1.

Quantitative real-time PCR

To estimate *SHOX* copy number (CN), a quantitative real-time PCR (qPCR) assay targeting exon 2 of the gene was designed (Applied Biosystem). Primers and probes were designed with the Primer Express v2.0 software (Applied Biosystems, Foster City, CA, USA) and tested for specificity using the NCBI's BLAST software. To estimate *SRY* CN, we used a commercially available qPCR assay (Roche, Mannheim, Germany). Reactions were performed in triplicate in a 10 µl final volume. The target gene (FAM-labeled) and the reference gene RNaseP (VIC-labeled) were co-amplified in a duplex qPCR. Four internal controls were always included in each experiment: (i) a subject with normal karyotype (calibrator); (ii) a subject with *SHOX* deletion; (iii) a 47,XXY subject and (iv) a no-template negative control. All runs were performed using 7900HT Fast System. Target genes CN was determined by relative quantification using the Copy-Caller-Software, v1.0 (Applied Biosystems, Foster City, CA, USA), based on the comparative ddCt method (Fig. 1). qPCR profiles are shown in [Supplementary data, Figs E–G](#). Assay reproducibility was validated by calculating the coefficient of variation (C_v) between five different experiments, in which CN was estimated for the three positive controls (normal man, deleted man and 47,XXY man). Standard deviations (S) and mean values (M) were calculated and C_v was computed as: $C_v = (S/M) \times 100$, denoting a very low inter-assay variability ranging between 0.4 and 2%.

Results

Array-CGH indicates that PAR-linked CNVs are mainly related to the individual's karyotype

Array-CGH was performed for twenty men with Yq microdeletions and twelve controls with an intact Y chromosome. Karyotype was available for 18 of the Yq microdeletion carriers, 13 with normal and 5 with abnormal karyotypes. All controls displayed a normal karyotype. This preliminary array-CGH analysis revealed that four of twenty carriers (20%) displayed CNVs, both losses and gains, at the PAR level. Of these four subjects, three carried an AZFb,c microdeletion and one carried a complete AZFc microdeletion. However, the array-CGH results were not unexpected since all the detected CNVs were explainable by the associated abnormal karyotype (Table 1A). None of the twelve controls displayed CNVs in the PARs. Validation of array-CGH-detected CNVs was performed using predesigned *SHOX* and *SRY* TaqMan CN Assays.

SHOX CN evaluation by qPCR

We screened another set of 204 men carrying various types of Yq microdeletions (including partial AZFc deletions and duplications) with *SHOX* qPCR alone, considering *SHOX* CN as a proxy of PAR1 rearrangements. Karyotype was available for 170 men, of which 164 had a normal karyotype and 6 had karyotype anomalies. We found that almost all subjects analyzed (201/204; 98.5%) had a normal *SHOX* CN, except for three men that displayed an extra copy of the *SHOX* gene (1.47%): samples D1056, P7806 (both carrying a complete AZFc microdeletion) and Mmp1000 (carrying an AZFbc microdeletion) (Table 1B). These patients were then screened for the *SRY* gene, as well. We found that one also had an extra copy of the *SRY* gene, whereas the other two had a normal *SRY* CN (Table 1B). Although Mmp1000's case history reported a 46,XYq-karyotype, the presence of two copies of the *SRY* gene clearly indicates that this is a 46,X,idel(Yp) with potential breakpoint in P6/P7 (Lange et al., 2009). No *SHOX* deletions were found in any of the Yq microdeletion carriers. If we only consider men with complete AZFc microdeletions,

Table 1 Patients with PAR-linked CNVs detected by qPCR and/or array-CGH.

A. Array-CGH analysis ^a						
Patient	Center	Deletion type	Tot sperm count (× 10 ⁶)	Karyotype	CNV type	
A416	Florence	AZFb,c	0	46,X,idic(Y)(q11.22)	Gain PAR1	
A372	Florence	AZFb,c	0	46,X,idic(Y)(q11.22)[31]/46,X,+mar[3]/45,X[16]	Loss PAR2	
A1389	Florence	AZFb,c	0	45,X[19]/46,X,del(Y)(q11.2)[81]	Loss PAR2	
A116 ^b	Florence	AZFc	0	45,X[90]/46,XY[10]	Loss PAR2	
B. qPCR screening ^c						
Patient	Center	Deletion type	Tot sperm count (× 10 ⁶)	Karyotype	SHOX CN	SRY CN
Mmp1000	Ancona	AZFb,c	0	46,XYq-	3	2
P7806	Padova	AZFc	0	46,XY	3	1
D1056	Muenster	AZFc	0	46,XY	3	1

CGH, comparative genomic hybridization; CN, copy number; CNV, copy number variation.

^aArray-CGH results were validated by subsequent qPCR.

^bFor patient A116, qPCR was not performed due to DNA sample extinction.

^cEach qPCR was performed in triplicate and each assay was repeated three times.

we observe that samples D1056 and P7806 are the only ones having *SHOX* CNVs, accounting for 1.31% (2/153) of carriers.

SHOX CNVs in microdeletions carriers and normal karyotype

If we solely consider men with Yq microdeletions and a normal karyotype, we observe that, on a total of 177 subjects (15 analyzed by array-CGH/qPCR and 164 analyzed by qPCR only), *SHOX* deletions were never found, whereas only *SHOX* duplications were detected in samples D1056 and P7806 (1.1%). Hence, 98.9% (175/177) of men with microdeletions and a normal karyotype did not display any *SHOX* CNVs. As a counterpart, we finally screened an extra set of 100 men, of two different nationalities (Italian and Spanish), displaying an intact Y chromosome and normal stature, and all resulted with a normal *SHOX* CN.

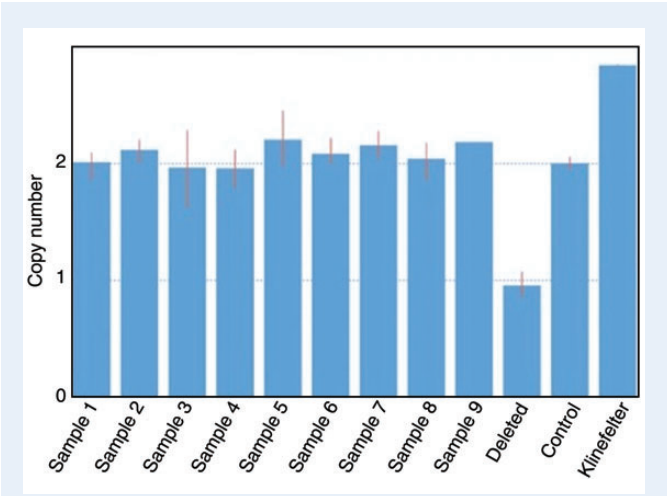


Figure 1 *SHOX* qPCR results. The figure represents data derived from a single experiment and elaborated by CopyCaller™ software v1.0. Deleted = 1 copy of *SHOX*; Control = 2 copies of *SHOX*; Klinefelter = 3 copies of *SHOX*.

SHOX duplications and the associated phenotype

A total of four patients, carrying both altered and normal karyotypes, displayed an extra copy of the *SHOX* gene. At the moment of consultation, all these patients presented with neither stature abnormalities nor any medical condition other than azoospermia (Table II). In the case of patient P7806, testicular sperm extraction upon bilateral TESE allowed the isolation of 13 fully matured spermatozoa from the right side and 0.025×10^6 from the left side. ICSI performed with the patient's cryopreserved spermatozoa resulted in the delivery of a healthy male child.

Discussion

Y-chromosome microdeletions represent the most frequent genetic cause of spermatogenic failure in infertile men, second only to Klinefelter syndrome (Krausz et al., 2011). In order to overcome their condition and father biological children, some carriers opt for treatment with ART. However, Yq genetic defects will be inevitably transmitted to their male offspring, who will predictably suffer from fertility problems. Therefore, genetic counseling for these couples is of inestimable importance.

Twenty years of interest in Y-chromosome microdeletions has produced a collection of numerous articles, but only a minority of these studies aimed to define whether Yq deletions might lead to other pathological conditions, beside spermatogenic failure. Microdeletions were reported in association with 45,X/46,XY mosaic karyotype and ambiguous genitalia (Papadimas et al., 2001; Patsalis et al., 2002, 2005; Papanikolaou et al., 2003; Tian et al., 2012); consistently, our data provide further evidence that Yq microdeletions do associate with mosaicism (as four samples presented 45,X cell lines), supporting the hypothesis that Y chromosomes bearing AZF deletions are more instable and thus predispose to the formation of Y-chromosome nullisomic cell lines.

Recently, a paper by Jorgez et al. (2011) reported that Y-chromosome microdeletions might not only cause spermatogenic failure but also increase the risk for PAR-related pathologies, especially emphasizing *SHOX* gene involvement. On a total of 87 men with Yq microdeletions, they found *SHOX* haploinsufficiency in five samples, four carrying a

Table II Phenotype characterization of Y-chromosome microdeletion carriers with an extra copy of *SHOX*.

Patient	D1056	P7806	A416	Mmp1000
Age	32	39	44	38
Height (cm)	180	188	n.a.	174
Nationality	German	Italian	Italian	Italian
Mean testis (cc)	7	6	n.a.	12
Tot sperm count (x10 ⁶)	0	0	0	0
Karyotype	46,XY	46,XY	46,X,idic(Y)	46,XYq-
FSH (UI/L)	29.9	21.9	17.9	13.9
LH (UI/L)	11.2	9.2	6.7	8.6
Testosterone (ng/ml)	3.77	3.54	5.75	4.61
Testis histology	Bilateral SCOS	Unilateral SCOS and hypospermatogenesis ^a	—	—
Other medical conditions	None	None	None	None

SCOS, Sertoli cell only syndrome; n.a., not available.
^aQuantitative alteration in which a very small portion of seminiferous tubules containing fully matured spermatozoa (13 in the right testis and 0.025×10^6 in the left testis) was found.

normal karyotype and one carrying an idic(Y); however, information is missing about these patients' phenotype. Moreover, they found nine men with *SHOX* duplications, all displaying abnormal karyotypes; still, stature was available only for six patients and clinical data only for four subjects, of whom only two displayed other medical conditions (congenital urological defects, diabetes and cataracts).

In this study, we provide the largest collection of men carrying Y-chromosome microdeletions analyzed so far in association with *SHOX* CN. A novelty of our study consisted in the inclusion of additional Y-chromosome rearrangements, such as gr/gr deletions and gr/gr deletion-duplications. Moreover, we included a very large number of men with complete AZFc microdeletion (4-fold higher than the previous study) because they represent the more plausible candidates for ART attempts. Our data seemingly indicate that *SHOX* CNVs are mainly linked to the individual's karyotype rather than the mere presence of microdeletions. As a matter of fact, of 177 deletion carriers with a normal karyotype only two (1.1%) displayed an abnormal *SHOX* CN: both patients bore duplications of the *SHOX* gene, but apart from a relatively high stature (though still within the normal range) the men's clinical work-up revealed no other pathological conditions than azoospermia. Literature is poor and unclear about *SHOX* over-dosage, which has been reported in association with normal to tall stature (Ogata *et al.*, 2001; Adamson *et al.*, 2002) as well as to more severe conditions such as LWD and idiopathic short stature (ISS) (Roos *et al.*, 2009; Thomas *et al.*, 2009; D'Haene *et al.*, 2010; Benito-Sanz *et al.*, 2012). In our study, the relatively low number of controls analyzed makes it difficult to define whether the *SHOX* duplications found in the two men with Yq microdeletions and a normal karyotype represent a polymorphism commonly found in the general population or are actually associated with the presence of the microdeletion. However, although we did not find *SHOX* over-dosage in controls, a recent study (Lopes *et al.*, 2013) reported the presence of a duplication burden on the Y chromosome, including PAR1, in men with no Yq microdeletions, suggesting that the 'duplication events' that we observed are unlikely to be related to the presence of microdeletions.

Conversely, a clear picture exists concerning *SHOX* haploinsufficiency and its deleterious effects on stature (Helena Mangs and Morris, 2007). In our screening, none of the men with microdeletions had *SHOX* deletions, contrasting the aforementioned hypothesis that microdeletion carriers are at a higher risk of developing pathologies caused by *SHOX* haploinsufficiency.

The reasons underlying the discrepancy between our study and the previous publication might be related to methodological issues, although the extremely succinct description of the method in the paper by Jorgez *et al.* (2011) renders it difficult to make any meaningful comparison. In our study, much care was taken to validate every step: template quality, inter-assay variability and all technical aspects were thoroughly addressed to avoid artifacts. Another potential bias might be ethnicity; but again, the ethnic background was not specified in the previous publication and thus comparison with our study cannot be done.

In summary, our study confirms the previously reported association between complete AZF microdeletions and 45,X/46,XY mosaicism. Moreover, we show that both partial and complete microdeletions in men with 46,XY karyotype are unlikely to be associated with *SHOX* haploinsufficiency, providing reassurance that the only established risk for ART offspring born from men with Yq microdeletions remains spermatogenic impairment.

Supplementary data

Supplementary data are available at <http://humrep.oxfordjournals.org/>.

Acknowledgements

We thank Marilena Pantaleo (Meyer Hospital, Florence, Italy), Elena Rossi (University of Pavia, Italy) and Albrecht Röpke (University of Münster, Germany) for technical support and helpful discussions. Special thanks to Esperança Martí, President of the Fundació Puigvert for her continuous support. We thank all the clinicians from the Andrology Unit of the University of Florence (M. Maggi, A. Magini and F. Lotti) and from the Fundació Puigvert (L. Bassas, O. Rajmil, J. Sarquella, A. Vives and J. Sanchez-Curbelo) who helped in providing samples for this study.

Authors' roles

All authors are justifiably credited with authorship, according to the authorship criteria. In detail, C.C.: design, a-CGH and qPCR analyses, acquisition of data, analysis and interpretation of data, drafting and revision of the manuscript; D.L.: partial AZFc deletion analysis; P.N.: qPCR in the 100 controls; S.V.: qPCR design; F.T., S.K., A.F., G.B., E.A., E.R.-C., G.F.: clinical definition of patients and DNA samples collection; S.G.: a-CGH analysis supervision; C.K.: conception and supervision of the project, funding, interpretation of data, drafting and revision of the manuscript and final approval given.

Funding

This work was supported by the Italian Ministry of University (grant PRIN 2010-2012 to C.K.), Tuscan Regional Health Research Program ('Progetto Salute 2009') to G.F., the Spanish Ministry of Health (grant FIS-11/02254) and the European Union 'Reprotrain' Marie Curie Network (project number: 289880 to C.K.).

Conflict of interest

None declared.

References

- Adamson KA, Cross I, Batch JA, Rappold GA, Glass IA, Ball SG. Trisomy of the short stature homeobox-containing gene (*SHOX*), resulting from a duplication-deletion of the X chromosome. *Clin Endocrinol* 2002; **56**:671–675.
- Benito-Sanz S, Royo JL, Barroso E, Paumard-Hernandez B, Barreda-Bonis AC, Liu P, Gracia R, Lupski JR, Campos-Barros A, Gomez-Skarmeta JL *et al.* Identification of the first recurrent PAR1 deletion in Leri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel *SHOX* enhancer. *J Med Genet* 2012; **49**:442–450.
- D'Haene B, Hellemans J, Craen M, De Schepper J, Devriendt K, Fryns JP, Keymolen K, Debals E, de Klein A, de Jong EM *et al.* Improved molecular diagnostics of idiopathic short stature and allied disorders: quantitative polymerase chain reaction-based copy number profiling of *SHOX* and pseudoautosomal region 1. *J Clin Endocrinol Metab* 2010; **95**: 3010–3018.

- Gervasini C, Grati FR, Lalatta F, Tabano S, Gentilin B, Colapietro P, De Toffol S, Frontino G, Motta F, Maitz S et al. SHOX duplications found in some cases with type I Mayer-Rokitansky-Kuster-Hauser syndrome. *Genet Med* 2010;**12**:634–640.
- Grigelioniene G, Schoumans J, Neumeyer L, Ivarsson A, Eklof O, Enkvist O, Tordai P, Fosdal I, Myhre AG, Westphal O et al. Analysis of short stature homeobox-containing gene (SHOX) and auxological phenotype in dyschondrosteosis and isolated Madelung deformity. *Hum Genet* 2001;**109**:551–558.
- Helena Mangs A, Morris BJ. The human pseudoautosomal region (PAR): origin, function and future. *Curr Genomics* 2007;**8**:129–136.
- Jorgez CJ, Weedon JW, Sahin A, Tannour-Louet M, Han S, Bourmat JC, Mielnik A, Cheung SW, Nangia AK, Schlegel PN et al. Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. *J Clin Endocrinol Metab* 2011;**96**:E674–E679.
- Krausz C, Chianese C, Giachini C, Guarducci E, Laface I, Forti G. The Y chromosome-linked copy number variations and male fertility. *J Endocrinol Invest* 2011;**34**:376–382.
- Lange J, Skaletsky H, van Daalen SK, Embry SL, Korver CM, Brown LG, Oates RD, Silber S, Repping S, Page DC. Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. *Cell* 2009;**138**:855–869.
- Lopes AM, Aston KI, Thompson E, Carvalho F, Goncalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A et al. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS Genet* 2013;**9**:e1003349.
- Ogata T, Matsuo N, Nishimura G. SHOX haploinsufficiency and overdosage: impact of gonadal function status. *J Med Genet* 2001;**38**:1–6.
- Papadimas J, Goulis DG, Giannouli C, Papanicolaou A, Tarlatzis B, Bontis JN. Ambiguous genitalia, 45,X/46,XY mosaic karyotype, and Y chromosome microdeletions in a 17-year-old man. *Fertil Steril* 2001;**76**:1261–1263.
- Papanicolaou AD, Goulis DG, Giannouli C, Gounioti C, Bontis JN, Papadimas J. Intratubular germ cell neoplasia in a man with ambiguous genitalia, 45,X/46,XY mosaic karyotype, and Y chromosome microdeletions. *Endocr Pathol* 2003;**14**:177–182.
- Patsalis PC, Sismani C, Quintana-Murci L, Taleb-Bekkouché F, Krausz C, McElreavey K. Effects of transmission of Y chromosome AZFc deletions. *Lancet* 2002;**360**:1222–1224.
- Patsalis PC, Skordis N, Sismani C, Kousoulidou L, Koumbaris G, Eftychi C, Stavrides G, Ioulianos A, Kitsiou-Tzeli S, Galla-Voumvouraki A et al. Identification of high frequency of Y chromosome deletions in patients with sex chromosome mosaicism and correlation with the clinical phenotype and Y-chromosome instability. *Am J Med Genet A* 2005;**135**:145–149.
- Roos L, Brondum Nielsen K, Tümer Z. A duplication encompassing the SHOX gene and the downstream evolutionarily conserved sequences. *Am J Med Genet A* 2009;**149A**:2900–2901.
- Simoni M, Bakker E, Krausz C. EAA/EMQN best practice guidelines for molecular diagnosis of y-chromosomal microdeletions. State of the art 2004. *Int J Androl* 2004;**27**:240–249.
- Tachdjian G, Aboura A, Portnoi MF, Pasquier M, Bourcigaux N, Simon T, Rousseau G, Finkel L, Benkhalifa M, Christin-Maitre S. Cryptic Xp duplication including the SHOX gene in a woman with 46,X, del(X)(q21.31) and premature ovarian failure. *Hum Reprod* 2008;**23**:222–226.
- Thomas NS, Harvey JF, Bunyan DJ, Rankin J, Grigelioniene G, Bruno DL, Tan TY, Tomkins S, Hastings R. Clinical and molecular characterization of duplications encompassing the human SHOX gene reveal a variable effect on stature. *Am J Med Genet A* 2009;**149A**:1407–1414.
- Tian L, Zhang JW, Shen CX, Du Y, Zhou X. Cytogenetics and Y chromosome AZF microdeletions in infertile patients with mosaic karyotype Klinefelter syndrome (46,XY/47,XXY/48, XXYY/49,XXXXY). *Zhonghua nan ke xue* 2012;**18**:545–550.

4.5. PAPER 5.

Novel insights into DNA methylation features in spermatozoa: stability and peculiarities.

Krausz C, Sandoval J, Sayols S, **Chianese C**, Giachini C, Heyn H, Esteller M.
PLoS One. 2012;7(10):e44479.

Novel Insights into DNA Methylation Features in Spermatozoa: Stability and Peculiarities

Csilla Krausz^{1,2,*}, Juan Sandoval^{3*}, Sergi Sayols³, Chiara Chianese¹, Claudia Giachini¹, Holger Heyn³, Manel Esteller^{3,4,5*}

1 Department of Clinical Physiopathology, Andrology Unit, University of Florence, Florence, Italy, **2** Fundacio Puigvert, Barcelona, Catalonia, Spain, **3** Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Catalonia, Spain, **4** Department of Physiological Sciences II, School of Medicine, University of Barcelona, Barcelona, Catalonia, Spain, **5** Institutio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

Abstract

Data about the entire sperm DNA methylome are limited to two sperm donors whereas studies dealing with a greater number of subjects focused only on a few genes or were based on low resolution arrays. This implies that information about what we can consider as a normal sperm DNA methylome and whether it is stable among different normozoospermic individuals is still missing. The definition of the DNA methylation profile of normozoospermic men, the entity of inter-individual variability and the epigenetic characterization of quality-fractioned sperm subpopulations in the same subject (intra-individual variability) are relevant for a better understanding of pathological conditions. We addressed these questions by using the high resolution Infinium 450K methylation array and compared normal sperm DNA methylomes against somatic and cancer cells. Our study, based on the largest number of subjects ($n = 8$) ever considered for such a large number of CpGs ($n = 487,517$), provided clear evidence for i) a highly conserved DNA methylation profile among normozoospermic subjects; ii) a stable sperm DNA methylation pattern in different quality-fractioned sperm populations of the same individual. The latter finding is particularly relevant if we consider that different quality fractioned sperm subpopulations show differences in their structural features, metabolic and genomic profiles. We demonstrate, for the first time, that DNA methylation in normozoospermic men remains highly uniform regardless the quality of sperm subpopulations. In addition, our analysis provided both confirmatory and novel data concerning the sperm DNA methylome, including its peculiar features in respect to somatic and cancer cells. Our description about a highly polarized sperm DNA methylation profile, the clearly distinct genomic and functional organization of hypo- versus hypermethylated loci as well as the association of histone-enriched hypomethylated loci with embryonic development, which we now extended also to hypomethylated piRNAs-linked genes, provides solid basis for future basic and clinical research.

Citation: Krausz C, Sandoval J, Sayols S, Chianese C, Giachini C, et al. (2012) Novel Insights into DNA Methylation Features in Spermatozoa: Stability and Peculiarities. PLoS ONE 7(10): e44479. doi:10.1371/journal.pone.0044479

Editor: Joel R. Drevet, Clermont-Ferrand Univ., France

Received: May 22, 2012; **Accepted:** August 7, 2012; **Published:** October 2, 2012

Copyright: © 2012 Krausz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the European Research Council (ERC) Advanced Grant EPINORC, the Ministerio de Ciencia e Innovación (MICINN) grant SAF2011-22803, the Health Department of the Catalan Government (Generalitat de Catalunya), the Italian Ministry of University (grant PRIN 2010-2012 to CK), the Spanish Ministry of Health (grant FIS -PI11/02254). JS is a “Juan de la Cierva” Researcher. ME is an Institutio Catalana de Recerca i Estudis Avançats (ICREA) Research Professor. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: c.krausz@dfc.unifi.it (CK); mesteller@idibell.cat (ME)

These authors contributed equally to this work.

Introduction

Human spermatogenesis is an outstandingly complex biological process which requires the concerted action of several thousands of genes [1]. An interesting feature of this biological process is the extremely large inter-individual variability of sperm production in healthy fertile men. The entity of this variation is well illustrated by a large recent study, reporting that total sperm number in the so called normal range (defined as 5th–95th percentile), varies from 40 millions to several hundred millions [2]. While a few genetic variants have been studied in relation to spermatogenic efficiency in normozoospermic men [3–6], the epigenetic aspects of such variations in the normozoospermic range is completely unexplored.

Apart from the large inter-individual variability of the above mentioned quantitative traits of spermatogenesis, semen of normozoospermic men contains a qualitatively (in terms of

motility and morphology) heterogeneous sperm population. With the advent and diffusion of assisted reproductive techniques, a number of sperm selection methods have been developed in order to obtain sperm subpopulations enriched with highly motile and morphologically normal spermatozoa to be used for *in vivo* or *in vitro* insemination. The rationale behind selection is mainly related to a predicted higher functional competency and a higher genomic integrity of selected spermatozoa. Interestingly enough, despite the same testicular environment, biochemical markers [7,8] as well as DNA integrity [9–12] show differences in distinct sperm fractions belonging to the same individual. It is still unknown whether these fractions also show differences in their methylation level.

Given that epigenetic signals such as DNA methylation and histone modifications are crucial for the proper functioning of the genome, phenotypic differences in sperm production (quantitative as well as qualitative traits) at both inter- and intra-individual level may also be due to an epigenetic variation. This hypothesis seems

to be plausible if we consider that the epigenome of mature spermatozoa mirrors a series of sequential epigenetic reprogramming events (demethylation and *de novo* methylation) which may generate substantial epigenetic variability. The sole study addressing the question about intra- and inter-individual DNA methylation changes in normozoospermic men was based on a 12,198-feature CpG island microarray [13]. The authors reported significant variations for 6 genes both at the intra- and inter-individual level, concluding that epigenetic variations may contribute to the variable semen phenotype. On the other hand, a limited inter-individual variability in DNA methylation was observed in two recent studies comparing two sperm donors by using a methylated DNA immunoprecipitation (MeDIP) procedure and promoter arrays [14] and a genome-wide shotgun bisulfite sequencing [15].

With respect to intra-individual variability of epigenetic marks in quality-fractioned sperm populations from normozoospermic and oligozoospermic men, data are available only for promoter CpG islands of two spermatogenesis candidate genes, *DAZ* and *DAZL* [16]. In this study, significant differences in the *DAZL* promoter methylation were observed between normal and defective germ cell fractions from the same individual. Other evidences for a potential association of DNA methylation defects and impaired sperm quality derives from studies based on the comparison of men with different sperm parameters including subjects with abnormally low sperm motility/morphology and sperm number [16–25].

Given the paucity of data on intra- and inter-individual variability of sperm DNA methylation, we aimed to provide a detailed description based on the analysis of a total of 487,317 CpG sites. Our first question was whether different quality-fractioned sperm populations deriving from the same individual displayed differences at the DNA methylation level i.e. whether “good” and “poor” quality spermatozoa differ not only in their metabolic markers and genome integrity but also in their methylation status. Our second aim was to assess the level of inter-individual variability by comparing the genome-wide methylation profiles of whole sperm populations and quality-fractioned sperm subpopulations of different normozoospermic subjects.

Finally, we aimed to get further insights into the sperm DNA methylome through the investigation on loci with “variable” and “conserved” DNA methylation levels between individuals and their relationship with chromatin modifications. In addition, in this part of the study, we focused on a singular topic, not addressed by others until now, that concerns the sperm methylation status of piRNAs (PIWI-interacting RNAs). This peculiar class of small non coding RNAs are specifically expressed in the testis and seem to be involved in the maintenance of genomic stability and germ cell function through the silencing, via DNA methylation, of mobile genetic elements such as transposons (reviewed in Aravin et al. [26]). In fact, knock-out mice models for the proteins involved in the piRNA biogenesis (MIWI, MILI, MIWI2) revealed a restoration of transposon activity, which is thought to be the cause of the observed sterility due to meiotic arrest [27,28]. However, given that piRNAs have recently been identified also in human cancer cells and somatic cells, it has been proposed that piRNAs regulate gene expression more broadly than previously predicted (for review see Juliano et al. [29] and Siddiqi and Matushansky [29,30]). In order to provide new insights into this largely unexplored topic, we investigated the piRNAs methylation status in spermatozoa and performed a comparative analysis with a differentiated somatic cell type (B cell) and a colorectal cancer cell line (HCT-116).

Our study, based on the largest genome-wide DNA methylation analysis available to date in a group of normozoospermic men, allowed us to both define the “normal” sperm DNA methylome with its peculiar features and discover a potential new role for sperm piRNAs in embryonic development.

Materials and Methods

Subjects

Ethics statement: All participating subject signed an informed consent and the project has been approved by the local Ethical Committee of the University Hospital Careggi.

Eight healthy normozoospermic individuals of Italian origin belonging to the upper normal range of sperm number were analyzed in this study. Sperm parameters, age and relevant phenotypic information are reported in Table S1. Considerable care was taken for the selection of subjects in order to provide a homogeneous group in terms of life style factors, age, BMI and semen characteristics. Special attention was paid in selecting only semen samples devoid of contaminating somatic cells in their ejaculate. The absence of leucocytes or uroepithelial cells was assessed by scoring 5 stained slides at the light microscope in all 8 samples. The purity of the swim-down fraction deriving from contaminating cells was documented by checking additional 5 slides at light microscopy. This procedure based on a two-step purity check granted a biologically irrelevant, if any, contamination in both whole semen and the swim-down fractions.

Three aliquots were obtained from each individual corresponding to: 1) whole sperm population after 1 hour from semen collection; 2) swim-up fraction; 3) swim-down fraction. For sample EC7, the swim-down fraction has been excluded due to DNA degradation. For 3 samples whole, semen at 2 Sperm DNA methylation profile largely hours (corresponding to the time at which the swim-up procedure ends) were also available for the comparison with the other fractions.

Sperm selection

Whole semen has been centrifuged on a 25% Percoll gradient (20 minutes) before the standard swim-up separation technique. Although much care was taken for the selection of samples in terms of lack of contaminating cells, this preliminary step further ensured the purity of the sperm population. The swim-up procedure allows spermatozoa with progressive motility to “swim up” into the culture medium while hypomotile/immotile spermatozoa remain behind. The upper fraction is denominated “Up”, whereas the fraction containing hypo/immotile spermatozoa is indicated in this manuscript as “Down or Dn”.

Sperm DNA extraction

Sperm DNA was extracted with an user-developed version of the QIAamp® DNeasy&Tissue Kit purification protocol. Fresh washed (in PBS) sperm was incubated 1:1 with a lysis buffer containing 20 mM TrisCl (pH 8), 20mM EDTA, 200 mM NaCl and 4% SDS, supplemented prior to use with 100 mM DTT and 250 ug/ml Proteinase K. Incubation was performed for 4 hours at 55°C with frequent vortexing. Prior to processing in the columns, 200 ul of absolute ethanol and 200 ul of the kit-provided lysis buffer were added to the samples. Then, purification was performed according to kit instructions.

Microarray-based DNA methylation analysis

DNA was quantified by Quant-iT™ PicoGreen dsDNA Reagent (Invitrogen) and the integrity was analyzed in a 1.3% agarose gel. Bisulfite conversion of 600 ng of each sample was

performed according to the manufacturer's recommendation for Illumina Infinium Assay. Effective bisulphite conversion was checked for three controls that were converted simultaneously with the samples. 4 µl of bisulfite converted DNA were used to hybridize on Infinium HumanMethylation 450 BeadChip, following Illumina Infinium HD Methylation protocol. Chip analysis was performed using Illumina HiScan SQ fluorescent scanner. The intensities of the images are extracted using GenomeStudio (2010.3) Methylation module (1.8.5) software. Methylation score of each CpG is represented as beta (β) value (β value <0.2 is considered as hypomethylated, >0.8 as hypermethylated). The 450K DNA Methylation array includes 485,764 cytosine positions of the human genome. From these cytosine sites, 482,421 positions (99.3%) are CpG dinucleotides, whilst only 3,343 sites (0.7%) correspond to CNG targets. Thus, from this point on we will use the term CpG, except when we refer specifically to putative CNG methylation. A general depiction of the 450K platform design, regarding functional genome distribution, CpG content and chromosome location, is reported in a previous validation study from our laboratory [31].

Data filtering

The 450K DNA methylation array by Illumina is an established, highly reproducible method for DNA methylation detection and has been validated in two independent laboratories [31,32].

Every *beta* value in the 450 K platform is accompanied by a detection p-value. We based filtering criteria on the basis of these p-values reported by the assay. We examined two aspects of filtering out probes and samples based on the detection p-values, selecting i) a threshold and ii) a cut-off. Previous analyses indicated that a threshold value of 0.01 allows a clear distinction to be made between reliable and unreliable *beta* values [31]. We selected the cut-off value as 10%. Following this criterion, we excluded all probes with detection p-values >0.01 in 10% or more of the samples and a total of 485,317 probes were included in the final analysis. We expect similar methylation level in neighbouring CpG sites given the strong correlation between CpG site methylation levels up to 150 bp.

Statistical analysis

In order to identify differentially methylated CpG sites between different quality fractioned sperm populations, a non parametric test (Wilcoxon rank sum test) has been performed. Linear regression coefficient has been calculated (Spearman's rho) both for intra and inter-individual variability of methylation levels. For all comparisons of methylation levels between different subgroups Fischer exact test was performed.

For the estimation of the degree of epigenetic dissimilarity between individuals we measured the Euclidean distances between two samples using the following equation:

$$d(a,b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

Where a_i and b_i represent the beta value for the i -essim CpG of samples "a" and "b", and "n" the number of CpG sites selected.

In addition, to estimate the inter-individual variability of the methylation status in the promoters of the 6 genes previously described as highly variable, we calculated for each gene a 100,000 permutations test with the distances of the three groups, in order to obtain a random distribution of possible mean distances and get a p-value for the mean variability among individuals in a group (the area below the distribution curve).

For the estimation of enrichment in biological processes we performed a hypergeometric test on biological processes defined by Gene Ontology [33].

Results

Comparison of genome-wide DNA methylation level in different quality-fractioned sperm populations deriving from eight normozoospermic men

The ejaculate of a normozoospermic man contains a qualitatively heterogeneous sperm population (in terms of different motility, morphology, metabolic and genomic features). This part of the analysis focused on intra-individual variation and addressed the biological question whether there are significant differences in methylation profiling between the "up" (enriched with highly motile and morphologically normal spermatozoa) versus "down" (poorly motile/immotile and morphologically abnormal spermatozoa) semen fractions in each subject.

Analysis of intra-individual variation in 485,317 CpG loci. By performing linear regression analysis, we observed an extremely high correlation (Pearson correlation coefficient ranging from $R^2 = 0.9896$ to $R^2 = 0.9982$) between "good" and "poor" quality sperm suspensions in all subjects (Table S2A). A representative example is given for sample EC01 in Figure 1. Accordingly, unsupervised hierarchical cluster analysis of the two tested groups was unable to cluster the "up" and "down" fractions into two distinct groups. Similarly, no significant differences were observed following comparison of the methylation levels in the 485,317 CpG sites between different sperm fractions ("up" versus "down", whole sperm population versus "down", whole sperm population versus "up"). By comparing epigenetic distances between the "up" and "down" fractions of the same individual, we found no significant differences except for one sample (EC12) with $p = 0.018$. Interestingly, this sample showed the lowest sperm count among the 8 normozoospermic individuals. Separately, we analyzed the intra-individual variation in selected CpG loci previously reported to be associated with poor sperm quality. To begin with, a few imprinted loci were analyzed in previous studies in relationship with a wide range of infertile phenotypes (oligozoospermia, oligoasthenozoospermia, asthenozoospermia). All previous studies reported methylation changes in a portion of infertile men, suggesting that impaired sperm production may be associated with methylation defects. A total of 2,386 CpGs belonging to 45 imprinted genes are present on the 450K array (24) and we analyzed the methylation status of their promoter regions in the "up" and "down" fractions. Similarly, we investigated on 289 CpGs belonging to 10 genes (*DAZ*, *DAZL*, *DAZAP*, *HRAS*, *KDM3A*, *MTHFR*, *NTF3*, *PAX8*, *RASGRF1*, *SEN*) showing DNA methylation changes in infertile men compared to normozoospermic controls as well as in different quality-fractioned sperm populations (such as *DAZ* and *DAZL*). In all cases, a homogeneous methylation pattern was observed in the two fractions derived from the same individual and, accordingly, the two sperm fractions derived from all analyzed subjects did not cluster separately (Figure 2 and 3).

Assessment of inter-individual variability in genome-wide DNA methylation profile. Although all subjects belonged to the upper normal range of sperm number, the whole semen fraction of each individual included a different proportion of "good" and "poor" quality spermatozoa. The most homogeneous sperm population containing the best quality spermatozoa was the "up" fraction. A slightly more pronounced inter-individual variability in DNA methylation profile has been observed compared to intra-individual variability between sperm fractions.

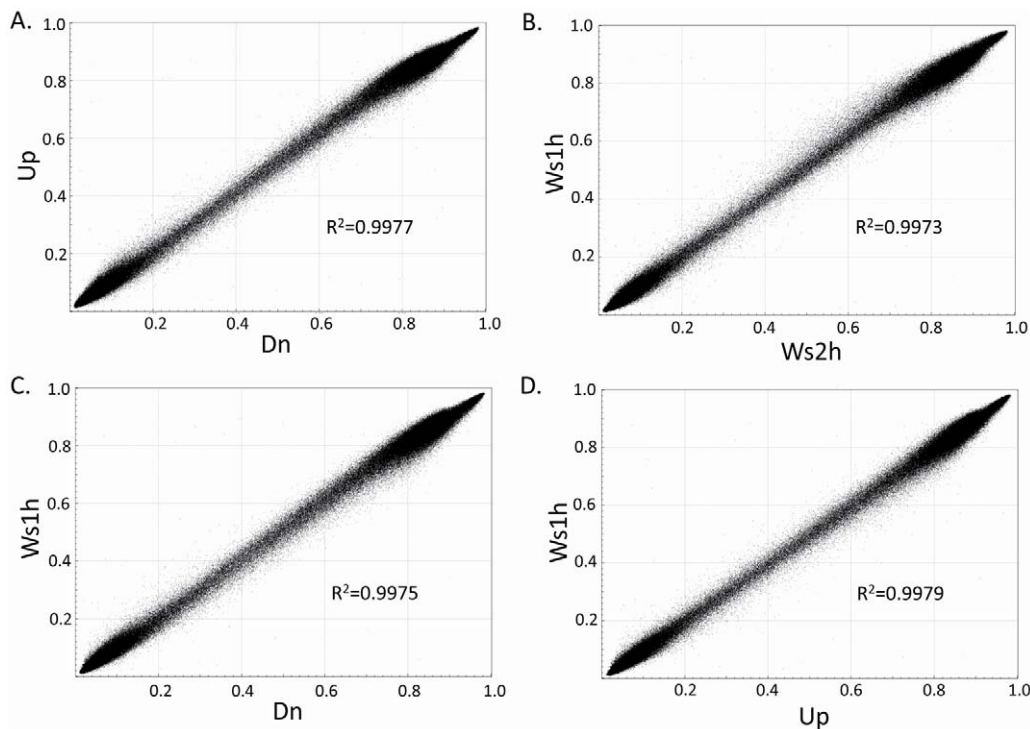


Figure 1. Scatter plots reporting CpGs methylation levels between different samples deriving from the swim-up sperm selection procedure in the same individual EC01: (A) swim-up (Up) sperm fraction versus swim-down (Dn) sperm fraction; (B) whole sperm population at 1h (Ws 1 h) versus whole sperm population at 2 h (Ws 2h); (C) whole sperm population at 1h versus swim-down sperm fraction; (D) whole sperm population at 1h versus swim-up sperm fraction. R^2 = Pearson coefficient.

doi:10.1371/journal.pone.0044479.g001

However, the linear regression coefficients were always >0.98 (Table S2B). A representative scatter plot comparing two individuals is shown in Figure 4 A–C.

In order to further explore inter-individual variability, we analyzed each type of sperm fraction using two additional approaches: i) we quantified the number of CpGs showing a standard deviation >0.2 in the methylation level (*beta* value) between individuals; ii) we measured the epigenetic distance (by the use of the Euclidean distance formula reported in materials and methods) between the methylation level of CpGs in different subjects. The number of differentially methylated loci i.e. showing a SD >0.2 between different subjects was 1,591 in the whole semen, 1,207 in the “down” fraction and 1,675 in the “up” fraction. This implies that for all comparisons the number of CpGs above the established threshold level was very low, representing $<0.3\%$ of all loci tested. The GO analysis of genes related to the 1,675 differentially methylated sites did not show any germ cell specific function. (data not shown).

By performing the comparison of DNA methylation distances across individuals considering all 485,317 CpG sites, a significantly higher variability has been observed in the swim-down sperm fraction ($p = 0.021$) in respect to the swim-up fraction (Figure 4D). However, it is worth noticing that the coefficient of variation is still very low in the swim-down fraction, e.g. 14% which indicates that the maximum epigenetic distance between individuals was limited to 45, that is significantly lower than the maximum distance possible e.g. 696 (see Table S3).

Assessment of methylation level in six gene promoters previously reported as having the highest intra and inter-individual variation. Significant DNA methylation variations have been reported for promoters of the following 6 genes: *BRCA1*,

BRCA2, *HTT (HD)*, *DMPK (DMI)*, *PSEN1*, *PSEN2* by Flanagan et al. [13]. In order to evaluate the entity of inter-individual variability, we calculated Euclidean distances for the beta values of the CpG sites in the above gene promoters among individuals of the three groups (Whole semen, “Down” and “Up”), plots are shown in Figure 5A–C. In addition, by performing permutation test of the epigenetic distances we found significant inter-individual differences for 4 out of 6 genes (*HTT*, *DMPK*, *PSEN1*, *PSEN2*) in the swim-down sperm fraction (p values: $2E-05$; 0.00096 ; 0.00348 ; 0.02 , respectively), while we observed no relevant variation in the swim-up fraction. In the whole sperm population sample, significance was reached only for *BRCA1* (0.01136).

Sperm genome-wide methylome description and its comparison with differentiated somatic cells

Sperm DNA methylation profile: general features. Given that the swim-up fraction, being enriched in the best quality spermatozoa, is the one used for assisted reproductive techniques, we aimed to provide a detailed description of genome wide DNA methylation profile of these cells. The average DNA methylation level of the 485,317 CpG sites was 45% (median value 35%). However, an interesting feature of the sperm DNA methylome is the polarization of DNA methylation level towards the two extremes: 86% of all markers are either severely hypomethylated ($<20\%$) or strongly hypermethylated ($>80\%$). Intermediate methylation level (20–80%) was observed only for 14% of CpGs. We defined, in each subject, the number of hypomethylated and hypermethylated loci for the whole genome as well as for the sex chromosomes and autosomes, separately (Table 1). The coefficient of variation of DNA methylation levels was minimal between subjects for the hypomethylated loci (0.9%) and slightly higher for

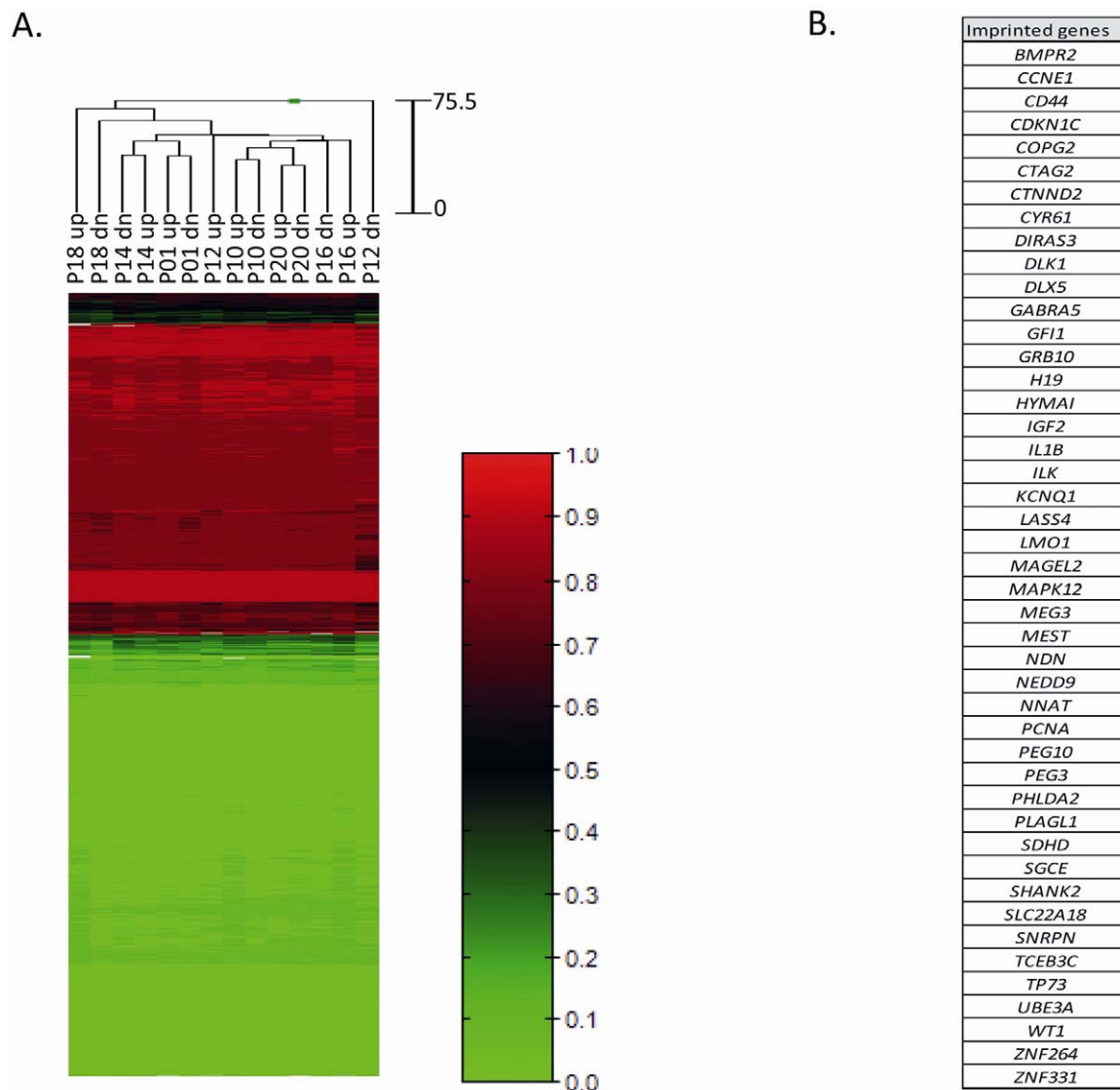


Figure 2. Heatmap displaying the methylation status of CpG loci ($n=2386$) mapping in the promoters of 45 imprinted genes in relation to quality-fractioned sperm populations (i.e. swim-up “up” and swim- “dn” sperm fractions). A) The dendrograms above the heatmap show hierarchical clustering based on the methylation data alone. Sperm populations and CpG loci are represented by columns and rows, respectively. Each cell indicates the CpG methylation level for one site in each sample. Methylation levels are represented in the scale on the right side of the heatmap and are referred lowest to highest as green (0.0) to red (1.0). (B) List of the 45 imprinted gene available in the array.
doi:10.1371/journal.pone.0044479.g002

hypermethylated loci (2.8%), suggesting a highly conserved profile both for hypo and hypermethylated loci. Accordingly, we found 95.8% of all hypomethylated loci to be conserved between individuals ($n=220,300$ CpGs), whereas in the hypermethylated loci the concordance was slightly lower, 86.1% ($n=161,542$ CpGs).

The separate analysis of autosomal (a total of 473,681 CpGs), X-linked (a total of 11,220 CpGs) and Y-linked CpGs (a total of 416 CpGs) revealed that X-linked loci are significantly more frequently hypomethylated than autosomal loci (64.5% versus 45.8%; $p<2.2 \times 10^{-16}$), as was the case also for the Y-linked loci (65.2% versus 45.8%, $p=3.458 \times 10^{-5}$) (Table 1). On the other hand, autosomal loci were significantly more frequently hypermethylated than X-linked loci (38.1% versus 21.6%; $p<2.2 \times 10^{-16}$). It is also worth noticing, that the highest percentage of

“conserved” hypomethylated loci was found for the X-linked loci (96.1%).

Sperm DNA methylation profile: comparative analysis of regions with conserved hypo/hyper methylation and differentially methylated loci between individuals. Subsequently, we identified loci displaying the same DNA methylation pattern in all subjects (“conserved” hypo or hyper) as well as loci showing different DNA methylation patterns (“variable” or “differentially methylated” loci). We analyzed the functional genomic distribution (promoter, body, 3’UTR, and intergenic), CpG content and neighborhood context. For the latter, we referred to: i) “island” as a DNA sequence (>200-bp window) with a GC content greater than 50% and an observed: expected CpG ratio of more than 0.6.; ii) “shore” as a sequence 0–2 Kb distant from the CpG island; iii) “shelf” as a sequence 2–4 Kb distant from the CpG island; iv) “open sea/

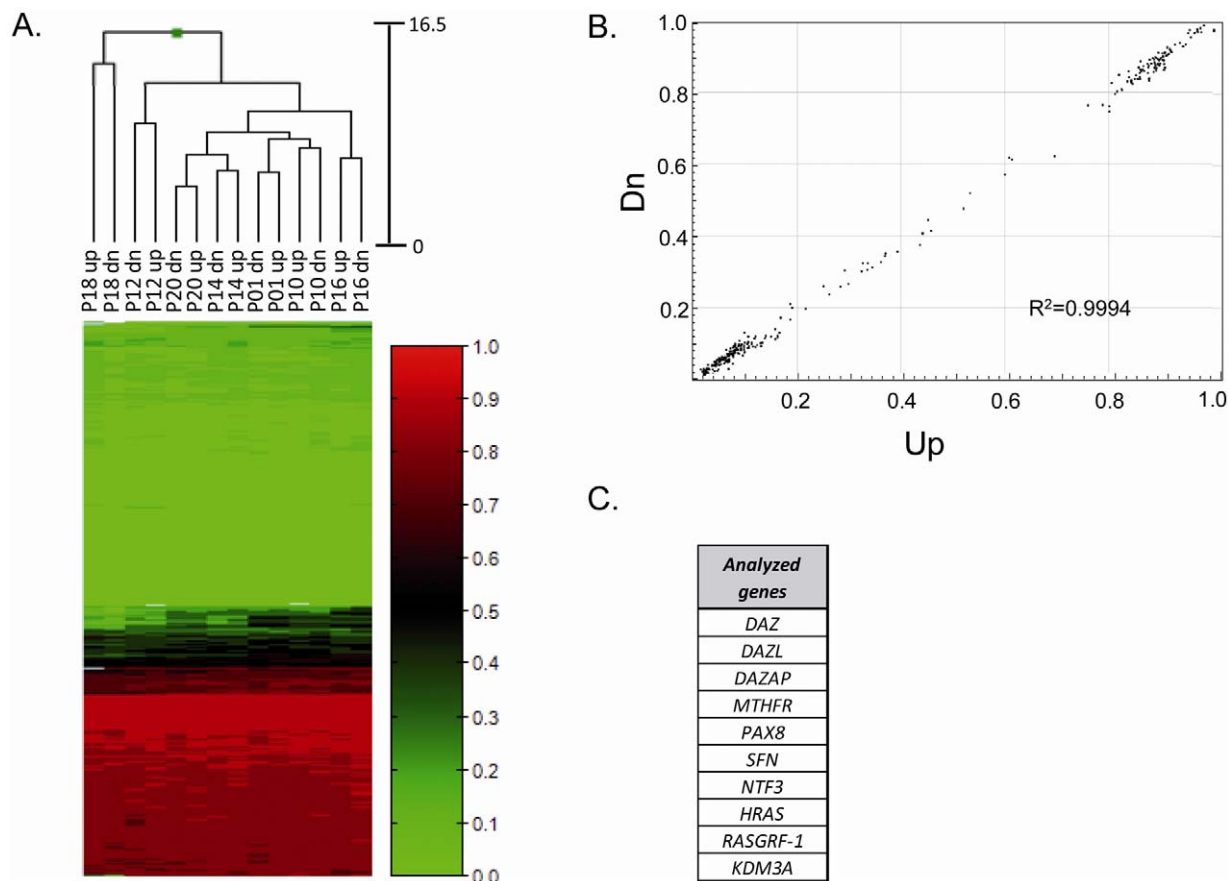


Figure 3. Heatmap displaying the methylation status of CpG loci (n=297) mapping in 10 selected genes in relation to quality-fractionated sperm populations (i.e. swim-up “up” and swim-down “dn” sperm fractions). A) The dendrograms above the heatmap show hierarchical clustering based on the methylation data alone. Sperm populations and CpG loci are represented by columns and rows, respectively. Each cell indicates the CpG methylation level for one site in each sample. Methylation levels are represented in the scale on the right side of the heatmap and are referred lowest to highest as green (0.0) to red (1.0). B) Scatter plot reporting CpGs methylation levels between quality-fractionated sperm populations (Up vs Dn) among different individuals. R^2 =Pearson coefficient. C) List of the 10 analyzed genes, selected because previously reported as differently methylated in infertile men compared to normozoospermic controls. doi:10.1371/journal.pone.0044479.g003

others” as the remaining sequence. The methylation categories were also analyzed in relationship with genomic locations related to RNA transcription (coding, non-coding and intergenic). Sharp differences were observed between the “conserved” hypo and hyper-methylated loci according to the functional genomic distribution as well as for the CpG and neighborhood context (Figure 6). Among all “conserved” hypomethylated loci 63.6% belonged to promoters, whereas this percentage was significantly lower in the hypermethylated loci, 19.5% ($p=4.14E-05$). A significant difference was observed also in the methylation status of gene body-linked CpGs which made up 47.7% of “conserved” hypermethylated loci and only a minority of the “conserved” hypomethylated loci (17.3%; $p=0.001357$). Given the high prevalence of promoters in the hypomethylated sites, almost 90% of the hypomethylated CpGs correspond to islands and shores (58.7% and 29.9%, respectively). On the contrary, 80% of “conserved” hypermethylated sites are either in the CpG poor island shelves or in “open sea” regions (18% and 63.3%, respectively).

Differentially methylated loci (defined as >0.2 standard deviation between individuals) have been identified only for 0.3% of all analyzed CpGs. Interestingly, the percentage of “variable” regions were lower in X-linked loci (0.2%) and were

completely absent in the 416 Y-linked loci. Intriguingly, the pattern of “variable” CpGs was more similar to the “conserved” hypermethylated loci than to the “conserved” hypomethylated ones, as a matter of fact the variation in DNA methylation between individuals is more pronounced in CpG-poor regions such as gene body, intergenic and “open sea” (Figure 6).

Gene Ontology analysis of “conserved” hypo and hypermethylated loci. Our next question was whether hypo and hypermethylated loci were linked to specific biological processes. By performing GO analysis, we found that the two methylation patterns are involved in completely distinct cellular processes (Table S4A). An outstandingly high association has been observed between hypomethylation and genes involved in metabolic and biosynthetic processes (among the first 20 significant associations, 11 are linked to metabolic and 5 to biosynthetic processes). On the contrary, hypermethylated sites, while associated with several different biological processes, did not show any association with metabolic and biosynthetic genes.

Analysis of DNA methylation levels in histone-enriched loci and gene ontology analysis. In a previous study, Hammoud et al.[14] defined the position of histone enriched loci in the sperm genome. We crossed our list of “conserved” hypo and hypermethylated loci with the list of histone positions referring to

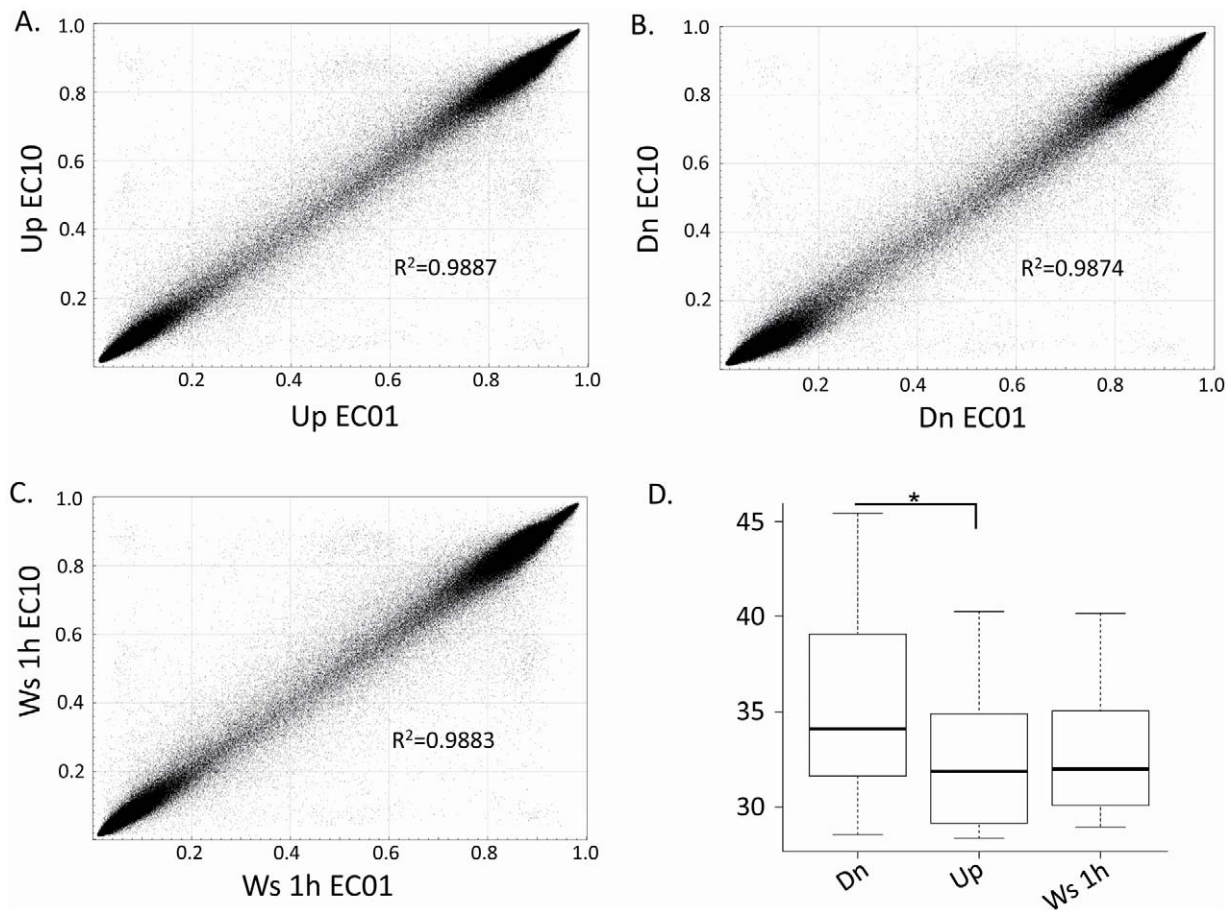


Figure 4. Representative scatter plots reporting CpG methylation levels between different individuals EC01 and EC10. (A) swim-up (Up) sperm samples; (B) swim-down (Dn) sperm samples; (C) whole sperm population at 1 h (Ws 1 h) samples; (D) Box plot representing the inter-individual variability of DNA methylation levels in total CpGs from the swim-up, swim-down and whole sperm population at 1h samples. The median value is shown. * corresponds to p value = 0.0213; R^2 = Pearson coefficient. The boxes describe the lower quartile (Q1, 25%), median (Q2, 50%) and the upper quartile (Q3, 75%); the whiskers extend 1.5 times the IQR from the box.
doi:10.1371/journal.pone.0044479.g004

the top 9,841 regions (FDR 40 cut off) and found a total of 30,591 CpGs in our array. The large majority (98.9 %) of these CpGs were hypomethylated ($n=30,244$) whereas only 1.1% of all histone-retained sites were hypermethylated ($n=347$). Similarly to the globally considered “conserved” hypo/hypermethylated sites, we observed sharp differences in the distribution according to functional genomic and CpG content criteria between hypo and hypermethylated loci enriched in histones, since promoters and islands are prevalent in the hypomethylated loci (74% and 90.5%, respectively), and scarcely represented in the hypermethylated loci (22.8% and 40.3%, respectively) (Figure 7). Hypermethylated loci at the global level (including both histone-enriched and histone-depleted i.e. protaminized segments) have been found mainly outside of the islands as well as shore and shelf areas and involve 63.3% of the so called “open sea/others” genomic regions whereas the same regions are present only in 21.6 % of hypermethylated histone-retained CpGs. Moreover, in hypermethylated loci overlapping with histones, islands were represented almost seven times more than in hypermethylated regions at the global level (40.3% versus 6.7%) (See Figure 6 and 7).

In agreement with Hammoud et al. [14] and Vavouri and Lenher [34], we also found that histone-retained hypomethylated regions were enriched with developmental genes (Table S4B) indicating that genes mapping to histone enriched regions are

related to completely distinct biological processes compared to the hypomethylated region at the global level (i.e. involved in metabolism). Concerning histone enriched hypermethylated regions the level of associations was much lower with specific biological processes (below $p < 10^{-4}$) and was more heterogeneous. An additional datum supporting the link between hypomethylation and histone retention of developmental genes was that among the 106 developmental genes available in the array, 62 presented in their promoters CpGs with $<20\%$ of methylation level as well as histone retention. On the contrary, no overlap with histones was observed in developmental gene promoters showing hypermethylation.

Comparison of the sperm DNA methylome with a differentiated somatic cell type. The average percentage of hypomethylated loci was significantly higher in sperm cells of the 8 subjects at the global, autosomal and sex-chromosomal levels ($p < 0.05$ for all comparisons) compared to the differentiated somatic cell (Figure 8). On the contrary, no differences were found concerning the percentage of hypermethylated sites. The most striking difference in hypomethylation concerned the X and Y chromosomes (Figure 8 and Table 1). Next, we searched for the number of equally methylated CpGs between spermatozoa and B cells. We found that 4% showed a differentially methylated pattern whereas 485,317 CpGs were equally methylated between the two

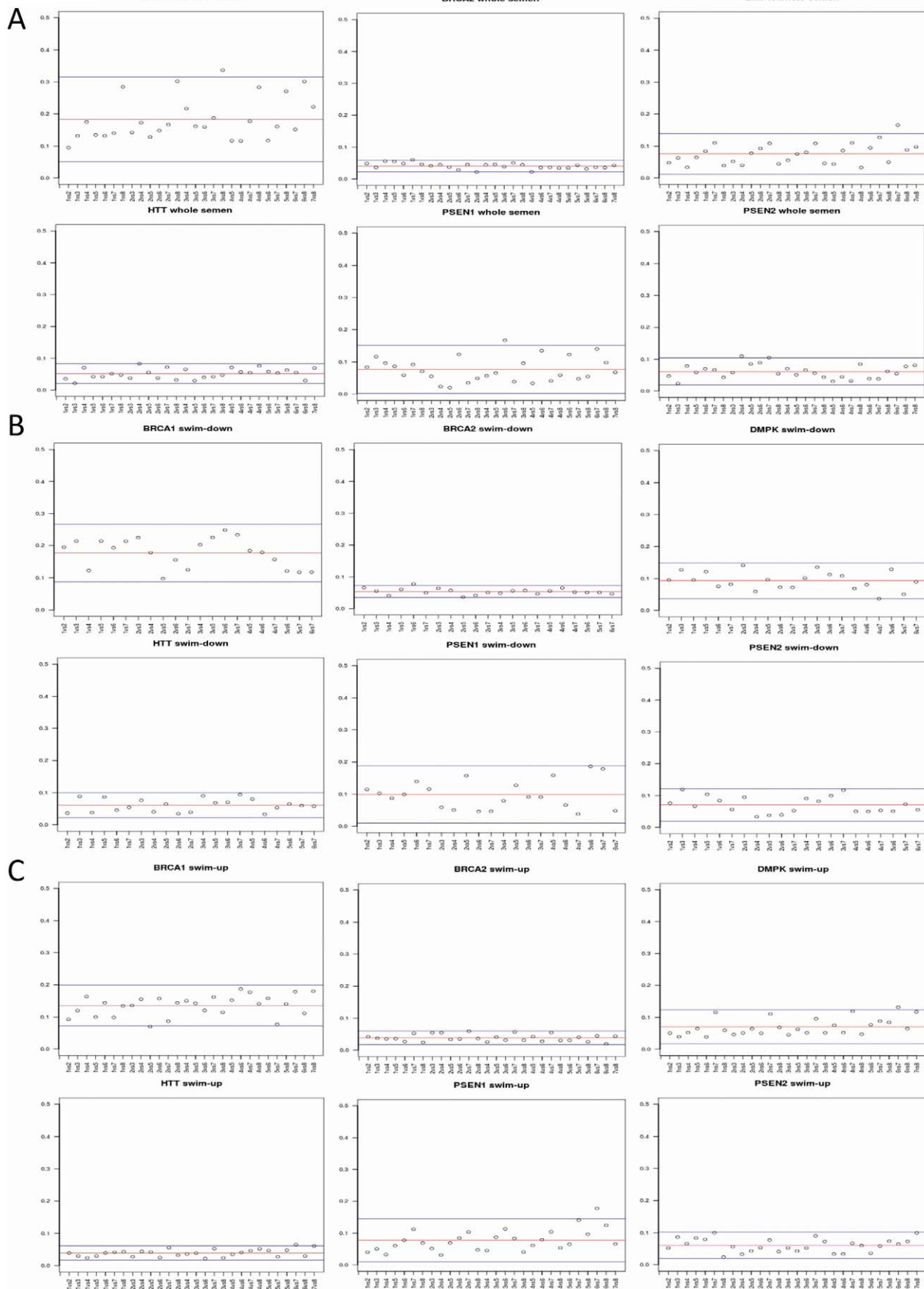


Figure 5. Intra-group epigenetic distances for the promoters of *BRCA1*, *BRCA2*, *HTT*, *DMPK1*, *PSEN1* and *PSEN2* genes. This distance represents the net dissimilarity of DNA methylation profiles between two sequences: the higher the distance, the more dissimilar are the compared samples. Different individuals were crossed with each other and Euclidean distances were calculated for beta-values of CpG sites as a marker of inter-individual variability in three different sperm subpopulations: A) Whole sperm population; B) swim-down and C) swim-up fractions. Numbers on the X axis indicate the identity of the pair-wise comparisons inside the experimental group: individuals EC01, EC07, EC10, EC12, EC14, EC16, EC18 and EC10 are numbered 1 to 8. Distance values are displayed on the Y axis. The top and bottom blue guidelines represent the 0.025 and 0.975 quartiles, while the red guideline represents the mean distance value.
doi:10.1371/journal.pone.0044479.g005

types of cells (96%). (Figure 9). Among the almost 20,000 CpGs showing a sperm-specific differentially methylated pattern (either hypo- or hypermethylation), those hypomethylated CpGs in spermatozoa, which were found to be hypermethylated in B cells, are the most represented proportion (76%). We analyzed the functional genomic distribution, CpG content and the associated RNA transcripts in the differentially methylated sites, separately for hypo- and hypermethylated sperm specific CpGs. The only significant difference consisted in the overrepresentation of “open sea/others” elements in the category of CpG content/neighborhood context of the sperm-specific hypermethylated CpGs compared to the sperm-specific hypomethylated sites (71% versus 30%, $p=0.00086$). The 15,799 CpGs showing a sperm-specific hypomethylated pattern included 6,140 gene promoter CpGs which belong to 3,344 distinct genes. The function of these genes is mainly related to metabolic and biochemical processes, and among the strongest associations resulted also DNA methylation involved in gamete generation and piRNA metabolic processes (Table S5A). Furthermore, we identified a total of 195 genes with sperm-specific hypomethylated gene promoters which were associated with histone-retained regions and involved in developmental processes (organogenesis, especially neuronal development) and in spermatogenesis (Table S5B).

piRNAs and DNA methylation status. We were interested in providing a detailed description of the methylation status of piRNA-linked CpGs in spermatozoa, B cells and cancer cells. To accomplish this purpose, we crossed the position of 15,000 piRNAs with unique positions in the genome present in the piRNABank (<http://pirnabank.ibab.ac.in/>) with the positions of the 487,517 CpGs available in the array. In order to include potential regulatory sequences, we extended the piRNA positions to ± 2 Kb and through these criteria a total of 2,591 unique piRNAs have been found to be covered by 7,528 CpGs on the array. In spermatozoa, 80% of the total array-available piRNA-linked CpGs ($n=6050$) revealed either very high or very low methylation levels. In fact, similarly to the global sperm DNA methylome, sperm piRNA-linked CpGs showed a sharply polarized methylation profile being 48.6% hypomethylated ($<20\%$ methylation level) and 31.8% hypermethylated ($>80\%$ methylation level). We found a significantly higher proportion of piRNA-linked CpGs within the total hypomethylated loci (3,657 out of 220,300) compared to those found within the hypermethylated loci (2,392 out of 161,542) ($p=1.585E-05$). In order to obtain more comprehensive characterization of sperm specific piRNAs-linked CpG methylation, we performed a comparative analysis with a differentiated somatic cell type (B cell) and a colon cancer cell type (HCT116).

By comparing the three cell types, we observed substantial differences in the methylation status of the piRNA-linked CpGs. In somatic cell, 95% of piRNA-linked CpGs show an intermediate methylation level, with a remaining 4.5% hypomethylated and 0.4% hypermethylated loci. On the contrary, similar to spermatozoa, cancer cells showed a polarization toward hypo/hypermethylation, but with an opposite pattern of methylation compared to spermatozoa i.e. 26% of the HCT116 cell piRNA-linked CpGs was hypomethylated and 53% was hypermethylated.

Next, we aimed to define the number of overlapping and distinct CpGs within the three cell types showing the same DNA methylation pattern (hypo or hypermethylation) (Figure 10). Sperm DNA methylation profile largely overlaps with that of the cancer cell, especially for the hypermethylated loci (86.8%). The overlap was 51.5% within the hypomethylated CpGs. On the contrary, there is only a limited number of overlapping CpGs with the somatic cell, with the largest overlap within hypomethylated loci (8.9%) and the lowest for the hypermethylated CpGs (1.1%). Given the functional importance of histone-retained regions in spermatozoa [14; 34], we extended our analysis to histone-retained regions associated to piRNAs. A total of 408 piRNA-linked CpGs revealed to be overlapping with histone-retained regions in spermatozoa and interestingly, 97% of them showed $<20\%$ of methylation level. When comparing the 342 hypomethylated piRNA-linked CpGs in B cells to the 3,657 hypomethylated piRNA-linked CpGs in the sperm, we found that all, except 16 CpGs, were also present in spermatozoa. However, when comparing the same 342 hypomethylated piRNA-linked CpG sites in B cells to the 408 hypomethylated histone enriched piRNA-linked CpG sites in the sperm, only 3.2% of them overlapped. The same phenomenon was observed for the cancer cell i.e. almost all piRNA-linked hypomethylated loci in the HCT116 cell (1883 out of 1959) overlapped with the 3657 hypomethylated piRNA-linked CpGs in the sperm whereas only 263/1959 were shared between the two cell types when comparing to the sperm histone-enriched loci (408 CpG sites).

We next focused on the characterization of sperm-specific piRNAs. Accordingly, we identified the sperm-specific hypo and hypermethylated sites i.e. not overlapping with any of the two other cell types. We performed a GO analysis for the genes overlapping sperm-specific piRNA in order to define the type of biological processes in which the associated genes are involved (Table S6). A total of 213 genes were identified in association with piRNAs showing exclusive hypomethylation in the spermatozoa. Strikingly, some of these genes are involved in embryonic development.

Discussion

The mammalian germ line undergoes extensive epigenetic reprogramming during development and gametogenesis. In pre-implantation embryo, a pattern of somatic-like DNA hypermethylation is established in all cells, including those which are destined to give origin to germ cells. This active *de novo* methylation process is followed by a widespread erasure of DNA methylation in primordial germ cells. Subsequently, another wave of *de novo* methylation takes place during spermatogenesis, ensuring a male germ line specific pattern of DNA methylation. The understanding of this complex process and the description of sperm DNA methylome have multiple implications, including evolutionary [15] and clinical aspects [35]. The entire sperm DNA methylome has been recently described by Molaro et al. [15] and it is based on the analysis of the whole semen (without quality fractioning) belonging to two sperm donors. Studies dealing with a larger group of subjects analyzed only a few genes or were based on low resolution arrays [16–19,21–25]. This implies

Table 1. Description of CpGs in terms of number and methylation status in the swim-up sperm fraction and in B cells.

Patients	autosomes						X chromosome						Y chromosome					
	total		hypo		hyper		hypo		hyper		hypo		hypo		hyper		hypo	
	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells
EC01	228576	166537	190746	186229	221114	161321	188198	182663	7193	5113	2477	3414	274	113	71	152		
EC07	231894	164595	191062	180334	224312	159359	188447	176829	7313	5098	2536	3351	276	110	79	154		
EC10	228642		191830		221172		189232		7204		2521		271		77			
EC12	230315		187601		222808		185109		7251		2414		269		77			
EC14	230458		187734		222904		185222		7287		2438		272		74			
EC16	226746		187086		219322		184649		7166		2362		265		75			
EC18	233778		175270		226164		172984		7342		2220		276		64			
EC20	228732		188955		221292		186421		7180		2458		268		76			
total CpG	485317				473681				11220				416					
mean	229893	165566	187536	183282	222386	160340	185032	179746	7242	5106	2428	3383	271	112	74	153		
sd	2209	1373	5263	4168	2142	1387	5163	4125	66	11	101	45	4	2	5	1		
percentage	47.4	34.1	38.6	37.8	46.9	33.8	39.1	37.9	64.5	45.5	21.6	30.1	65.2	26.8	17.8	36.8		

Patients	"conserved" (total)						"conserved" (autosomes)						"conserved" (X chromosome)						"conserved" (Y chromosome)					
	hypo		hyper		hypo		hypo		hyper		hypo		hypo		hyper		hypo		hypo		hyper		hypo	
	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells	sperm	B cells
total CpG	220300	160674	161542	173119	213088	155602	159532	169793	6958	4972	1953	3185	256	102	58	142								
percentage	95.8	97.0	86.1	94.5	95.8	97.0	86.2	94.5	96.1	97.4	80.4	94.2	94.3	91.5	78.2	92.8								

The number of hypomethylated and hypermethylated loci are indicated as total, autosomic, X chromosome and Y chromosome-linked. A) Sperm CpG numbers refer singularly to the eight normozoospermic men, while B cells CpG numbers belong to two different subjects. The mean CpG number \pm DS and the percentage calculated in respect to the mean total CpG number for each group are reported in the middle panel; B) Number of CpGs conserved among individuals: the first raw reports the number of CpGs shared by individuals, while the second raw reports the percentage of conserved CpGs compared to the mean CpG number reported in panel A.

doi:10.1371/journal.pone.0044479.t001

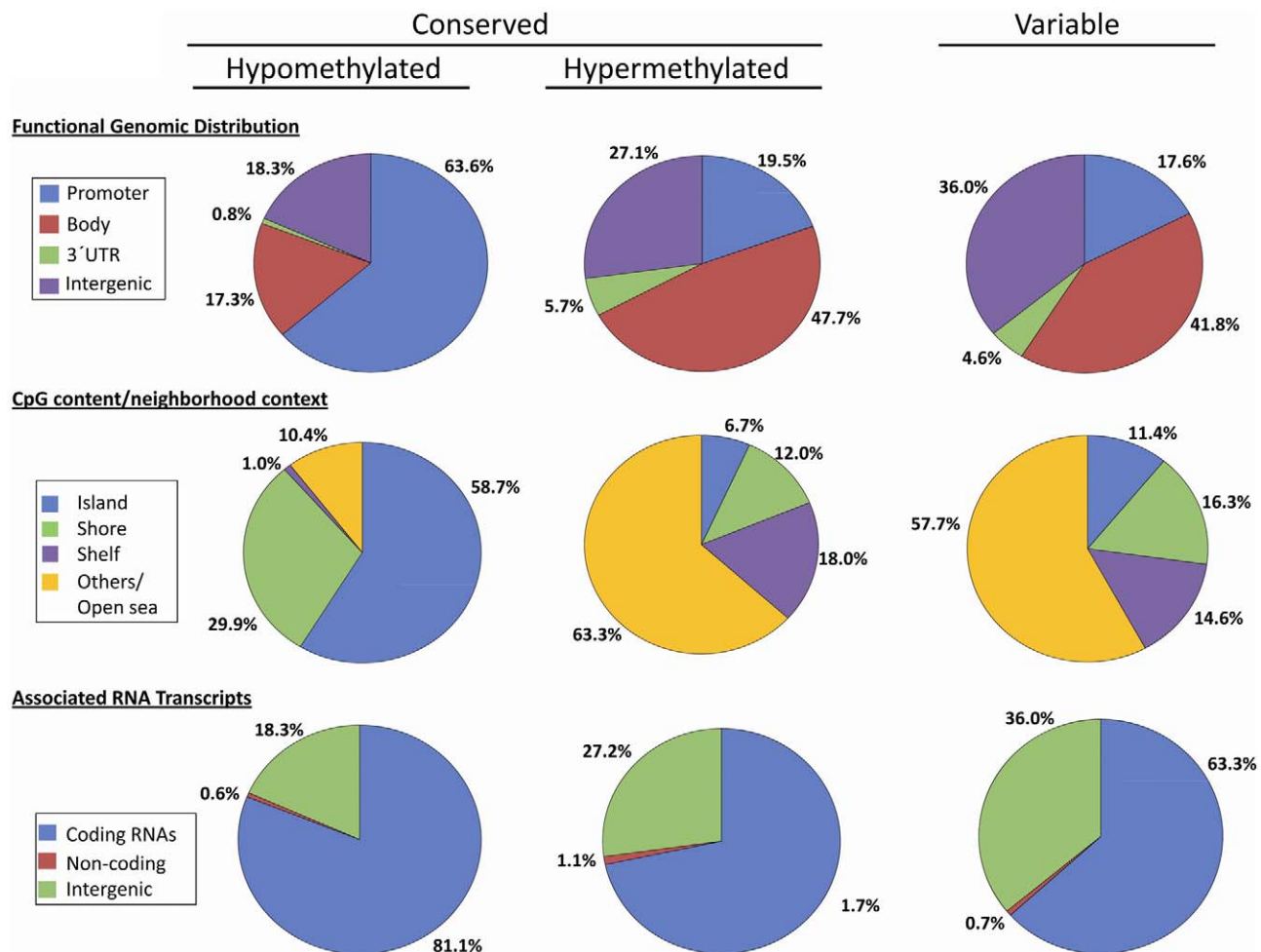


Figure 6. Sperm DNA methylation profile (swim-up sperm samples) according to i) functional genomic distribution; ii) CpG content/ neighbourhood context and iii) associated RNA transcripts. (A) Distribution of hypomethylated ($n = 220,300$) and hypermethylated ($n = 161,542$) CpGs conserved within individuals B) Distribution of variable CpGs within individuals ($n = 1674$). doi:10.1371/journal.pone.0044479.g006

that information about what we can consider as a normal sperm DNA methylome and whether this methylome is stable among different normozoospermic individuals is still missing. We addressed the above questions by using the 450 K platform which allowed us to provide the most extensive and comprehensive investigation on DNA methylation profile, available to date, on quality fractioned sperm populations in a group of normozoospermic subjects. By comparing data from the whole sperm DNA methylome [15] with that obtained with our array, we found a highly significant correlation $Rho = 0.97$ (Figure S1), indicating that our data and conclusions are highly reliable.

Our first aim was to provide data on sperm DNA methylation profile in quality-fractioned spermatozoa from the same subjects. Human semen is peculiar for the heterogeneity of its sperm population presenting a number of different qualitative features that include kinetic, morphological, metabolic and genetic/ chromatin differences. It is for this reason that sperm selection methods have been developed for assisted reproductive techniques in order to obtain a highly enriched subpopulation of spermatozoa exhibiting the best structural and functional characteristics, indicative of optimal fertilizing ability. The question whether these quality differences between sperm subpopulations are also reflecting modifications in the DNA methylation pattern has not

been addressed so far. In fact, all studies published to date, except for one, focused on either whole semen or just one selected sperm subpopulation. Our analysis of 487,517 CpGs revealed a profound stability of the sperm DNA methylome without significant differences between sperm fractions enriched in “poor” (swim-down) and “good” quality (swim-up) spermatozoa. For all comparisons we obtained surprisingly high correlations ($R^2 > 0.989$) and the two subpopulations did not show distinct clustering of their methylation profiles. However, by analysing the epigenetic distances between the two sperm fractions we were able to detect a significant difference only in one subject, although the correlation between his two sperm subpopulations was high also in this case, $R^2 = 0.9896$. We separately analyzed a series of genes for which DNA methylation defects had been previously reported in association with impaired sperm production/quality that included 45 imprinted genes, available on the array, and 10 additional genes selected from the literature. Despite expectation, the DNA methylation profile of these genes showed no differences in relationship with sperm quality. These data indicate that in normozoospermic men, the global DNA methylation profile is not affected significantly by structural and functional differences between sperm subpopulations. The extensive conservation of the DNA methylation status is especially surprising if we consider

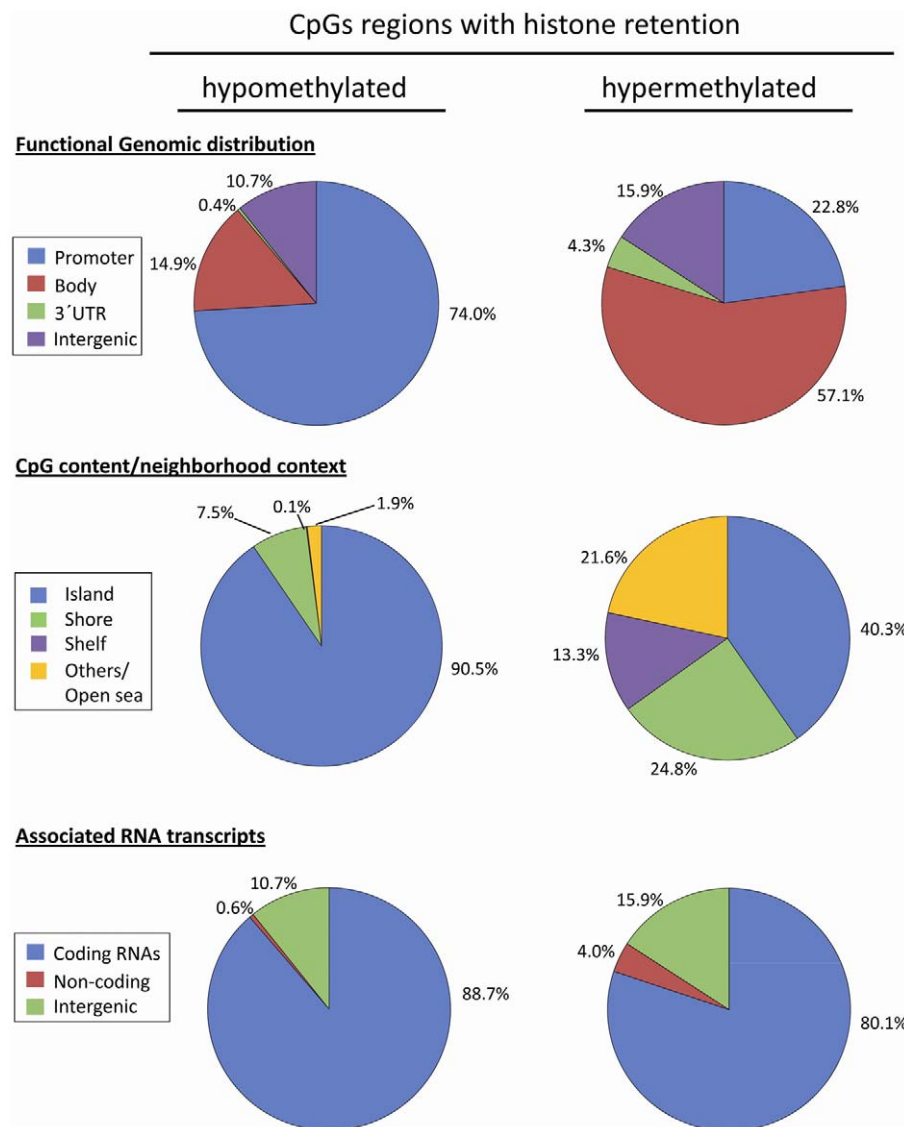


Figure 7. Sperm DNA methylation profile in histone-enriched regions according to i) functional genomic distribution; ii) CpG content/neighbourhood context; iii) associated RNA transcription. (A) Distribution of hypomethylated ($n=347$) CpGs in swim-up sperm samples. (B) Distribution of hypermethylated ($n=30244$) CpGs in swim-up sperm samples.
doi:10.1371/journal.pone.0044479.g007

that differences have been described also at the metabolic level of “poor” quality spermatozoa which could theoretically influence the DNA methylation process [7,8].

The definition of sperm DNA methylation profile between different normozoospermic subjects derives from a previous observation showing a significant epigenetic variability in human germ cells. Our aim was to further explore this observation both by increasing the number of analyzed CpGs (the previous study analyzed only 12,198 CpG sites) and by comparing different sperm fractions from different normozoospermic individuals. Our data, clearly proved that the methylation pattern in different individuals showing similar sperm characteristics without contaminating cells is highly conserved. In fact, the discrepancy with the previous study may well be due to a technical issue i.e. to the presence of contaminating somatic cells, which could account for the observed inter-individual differences in the methylation profile. The highest correlation was found in the selected fraction enriched with “good” quality spermatozoa with $R^2 > 0.98$. This observation

indicates that regardless of slight differences in life style factors, age and BMI, those cells which are designated to the fertilization process (good quality sperm-enriched fraction) show a highly stable sperm methylation profile between individuals. The few moderate smokers (< 10 cigarettes/day) included in the study did not cluster together, however it remains an important question whether sperm methylome can be altered by heavy smoking or other exogenous factors.

For the general description and comparative analyses of the sperm DNA methylome, we focused on the fraction enriched with “good” quality spermatozoa showing a complete lack of significant inter-individual differences. An interesting feature of the “normal” sperm DNA methylome is its highly polarized methylation profile towards the two extreme of DNA methylation levels: hypomethylation ($< 20\%$) and hypermethylation ($> 80\%$). We found that 96% of all CpGs belonged to one of the above categories. Hypo- and hypermethylated loci were highly conserved in different individuals reaching to a concordance of 95% for hypomethylated CpGs

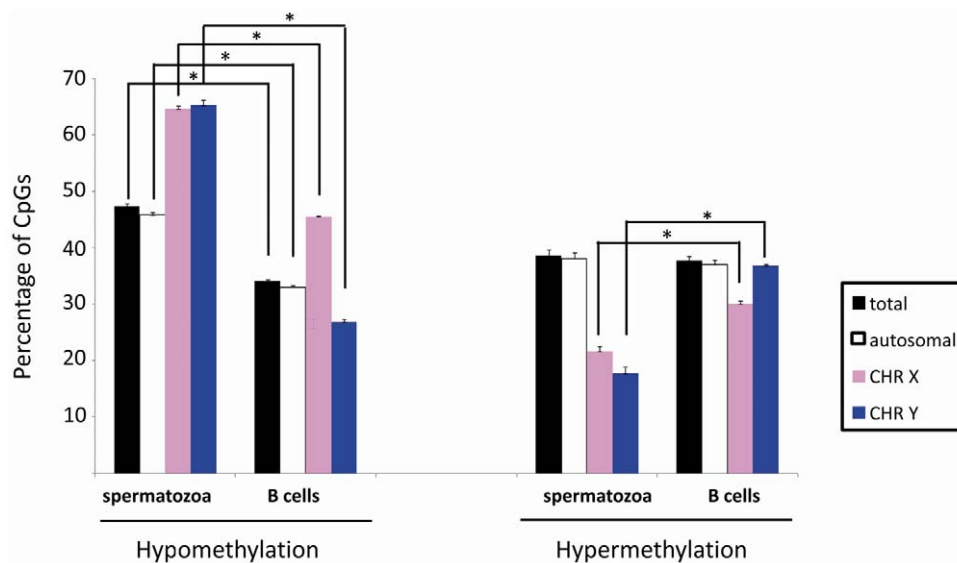


Figure 8. Bar graph illustrating the percentage of hypermethylated and hypomethylated CpGs in swim-up sperm samples and B cells. The number of detected CpGs varies according to the data extrapolation performed separately for each tested group: i) total CpGs, ii) autosomal CpGs, iii) X chromosome-linked CpGs and iv) Y chromosome-linked CpGs. * corresponds to p values <0.05 (the whiskers show the SD; n=7).

doi:10.1371/journal.pone.0044479.g008

and to 83.3% for the hypermethylated ones. These, so called “conserved CpGs” were further analyzed in comparison with the relatively few “variable CpGs” (0.3%) present in spermatozoa and with the B cell DNA methylation profile. The qualitative analysis of hypo-, hyper- and variably methylated regions showed significant differences between the conserved hypomethylated loci and the other two methylation categories; in fact, we observed a significant enrichment with promoters (63.6%) and islands/shores-linked CpGs in the hypomethylated loci. On the other hand, among the hypermethylated and “variable” CpGs there was a significant overrepresentation of gene body-linked CpGs which, together with intergenic CpGs, build up >60% of all CpGs. The high inter-individual conservation of hypomethylated loci, especially abundant in promoter regions, suggests that normal spermatogenesis requires strictly controlled methylation levels in specific gene promoters. At this regard, for the first time we provide evidence about an exceptionally high number of “conserved” hypomethylated X and Y chromosome-linked loci which further supports previous predictions on the importance of sex chromosome linked genes in spermatogenesis and stimulates further research on the sex chromosome methylation status in pathological conditions [36]. On the other hand, “variable” loci mainly in gene bodies and intergenic sequences may indicate their irrelevant role in spermatogenesis or may represent epigenetic changes which may act as fine-tuners of spermatogenetic efficiency and thus may contribute to the inter-individual variability of sperm production in normal healthy men.

An increasing number of studies are converging on the importance of histone-retained regions in spermatozoa for embryo development. The first study by Hammoud et al. [14] posited that genes involved in early embryonic development had a distinct chromatin status in sperm, being hypomethylated, histone-retained, enriched in H3K4me3 marks, and thus poised for expression. On the other hand, Brykczynska et al demonstrated that histone-retained regions with H3K27me3 mark may also play a role in post-fertilization, whereas histone H3Lys4 demethylation (H3K4m2) marks genes which are relevant in spermatogenesis

[37]. In addition, an other recent study reports a striking link between the retention of nucleosomes in sperm and the establishment of DNA methylation-free regions in the early embryo [34]. By using the 450 K array, we found that “conserved” hypomethylated CpGs mapping inside histone-enriched regions were associated with genes involved in developmental processes. Accordingly, the majority of developmental gene promoters available in the array were mapping inside histone-retained regions. Interestingly, the correlation with developmental genes was missing when the entire set of “conserved” hypomethylated regions were analyzed. In fact, genes belonging to this category are, indeed, involved in metabolic processes which indicate a differential biological function of genes situated in histone-enriched and histone-depleted regions.

The most relevant finding concerning the comparison between the DNA methylation profiles of the male germ cell and the B cell, is that only a minority of CpGs showed differential methylation (4.6%) between the two cell types and was mainly due to the overrepresentation of hypomethylated loci in spermatozoa. A total of 3,344 distinct genes were related to sperm-specific hypomethylated CpGs and among the strongest associations appeared “DNA methylation involved in gamete generation” and “piRNA metabolic processes”. Similarly to the general sperm DNA methylome data, those genes (n = 195) which were hypomethylated in histone-retained regions were involved in developmental processes (organogenesis, especially neuronal development) and spermatogenesis. The different methylation, in respect to the somatic cell, of the promoters of spermatogenesis genes is in accordance with the well known importance of epigenetic regulation of cell specific functions. The association with developmental genes further reinforces the hypothesis about a programmed histone retention in spermatozoa, which would serve for rapid activation of genes involved in embryonic development.

Finally, the complete lack of studies focusing on the methylation status of piRNAs in spermatozoa prompted us to provide a detailed analysis of this specific class of small non coding RNAs. Although piRNAs were first described as specifically expressed in the testis,

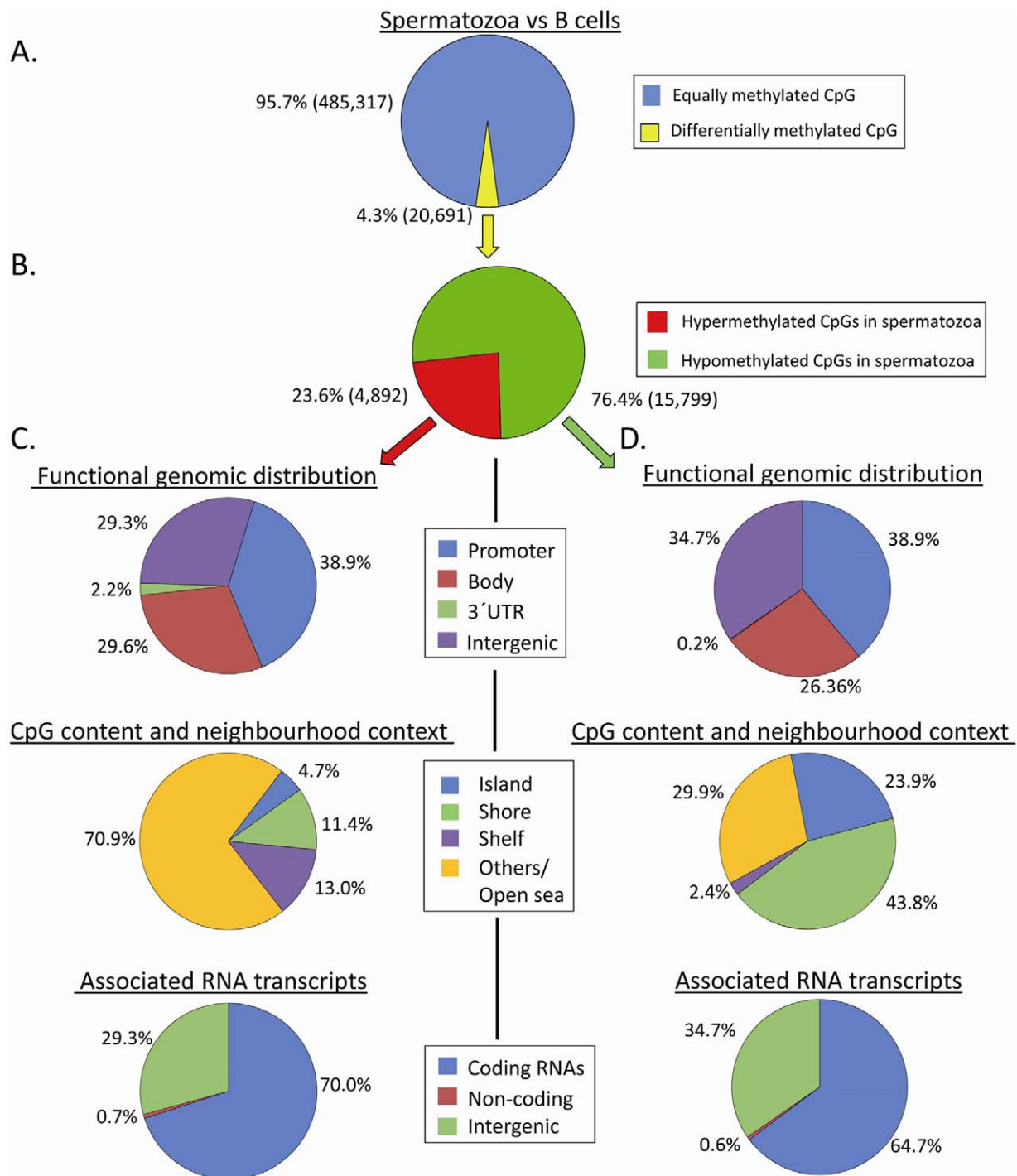


Figure 9. Spermatozoa versus B cell: a 450K DNA methylation portrait. (A) Graph showing percentages of equally and differentially methylated CpG sites in swim-up sperm samples compared to B cells. (B) Graph showing percentages of hypermethylation and hypomethylation in spermatozoa relating to the differentially methylated CpGs proportion (4.3%). Graphs describing the hypermethylated (C) and hypomethylated (D) sites according to their i) functional genomic distribution; ii) CpG content/neighborhood context and iii) association with RNA transcripts.
doi:10.1371/journal.pone.0044479.g009

recent data suggest their potential role in tumorigenesis and in somatic cell function [29,30]. In addition the presence of piRNAs has been also described in spermatozoa [38]. The 450K array is able to provide the characterization of a total of 2,591 unique piRNAs covered by 7,528 CpGs on the array. In spermatozoa we found a significantly higher proportion of

piRNA-linked CpGs within the total hypomethylated loci compared to those found within the hypermethylated loci ($p = 1.585 \times 10^{-5}$). The preferential hypomethylation of piRNAs was evident also in comparison with two other cell types: a differentiated somatic cell type (B cell) and a colon cancer cell type (HCT116). In fact, in spermatozoa 48.6% of CpGs were

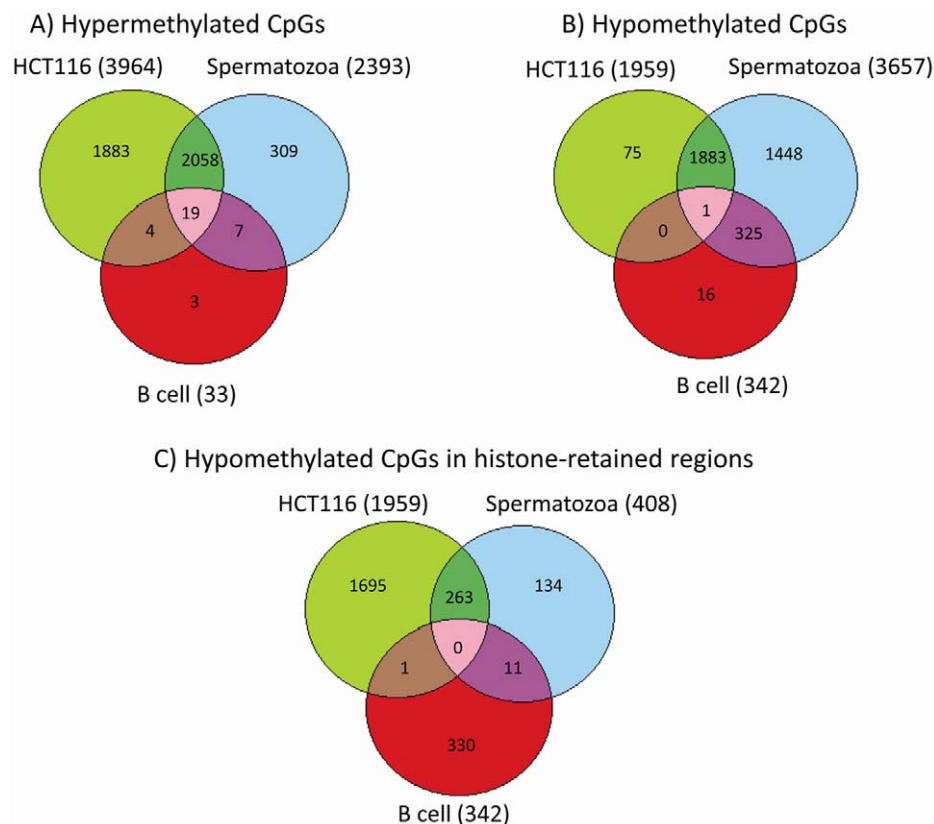


Figure 10. Venn diagram reporting unique and overlapping CpGs in/between the three cell types. A) Hypermethylated CpGs. B) Hypomethylated CpGs. C) Hypomethylated CpGs in histone-retained regions. HCT116: colorectal cancer cell line. doi:10.1371/journal.pone.0044479.g010

hypomethylated, whereas the percentages were 26% and 4.5% in HCT116 and B cell, respectively. Intriguingly, among those piRNAs which were located in histone-retained regions in spermatozoa, 97% of them showed low level of methylation. This observation represents a starting point for future studies aimed to explore the biological significance of these cell-dependent differences. An additional novel finding concerns the involvement of piRNA-related genes in distinct biological processes according to the methylation status of the related piRNAs. Most importantly, hypomethylated piRNAs are linked to genes associated with embryonic development and cell adhesion. Interestingly, piRNAs in histone-retained regions, showing hypomethylation exclusively in spermatozoa, are involved in the negative regulation of metabolic and biosynthetic processes which could be potentially relevant to the embryo. Given that almost all of these piRNAs are located inside or in the 3'UTR regions of the abovementioned genes, a potential RNA interfering mechanism can be hypothesized [29,39]. The interference with those RNAs which would have a negative regulatory effect on metabolism and biosynthesis may have an important biological function in early embryonic development.

In conclusion, our study, based on the largest number of subjects ever considered for such a high number of CpGs, provided clear evidence of a highly conserved DNA methylation profile among normozoospermic subjects. We also demonstrated that sperm methylation is stable in different quality-fractioned sperm subpopulations of the same individual i.e. sperm methylation is not altered in “poor” quality spermatozoa of normozoospermic men despite the fact that these cells are clearly different from a metabolic and DNA integrity point of view. In addition, our array-based analysis provided both confirmatory and novel data concerning the “normal” sperm DNA methylome, including its peculiar features in respect to somatic and cancer cells. Our description about a highly polarized sperm methylation profile, the clearly distinct genomic and functional organization of hypo versus hypermethylated loci and the association of histone-enriched hypomethylated loci with embryonic development, which we now extended also to hypomethylated piRNAs-linked genes, represents a solid basis for future basic and clinically oriented research.

spermic men despite the fact that these cells are clearly different from a metabolic and DNA integrity point of view. In addition, our array-based analysis provided both confirmatory and novel data concerning the “normal” sperm DNA methylome, including its peculiar features in respect to somatic and cancer cells. Our description about a highly polarized sperm methylation profile, the clearly distinct genomic and functional organization of hypo versus hypermethylated loci and the association of histone-enriched hypomethylated loci with embryonic development, which we now extended also to hypomethylated piRNAs-linked genes, represents a solid basis for future basic and clinically oriented research.

Supporting Information

Figure S1 Comparison of the methylation levels obtained in the array versus data reported in Molaro et al paper (ref 15). Heatmap generated from the distance correlation matrix for the 8 individual samples “up” (A) and “down” (B), the scale of the correlation is shown above the matrix (scale values from 0 to 0.03); Correlation scatter plots between Molaro's data vs the average methylation level for all “up” samples (C) and “down” samples (D), the Pearson correlation coefficient (ρ) is shown. (DOC)

Table S1 Clinical description of the 8 normozoospermic individuals. (DOC)

Table S2 Analysis of intra-and inter-individual variability of the sperm DNA methylation profile: estimate of correlation coefficients.

(DOC)

Table S3 Analysis of inter-individual variability of the sperm DNA methylation profile: epigenetic distance and coefficient of variation.

(DOC)

Table S4 Biological processes associated with genes linked to “conserved” hypo- and hypermethylated CpG loci in spermatozoa.

(DOC)

Table S5 Biological processes associated with genes linked to sperm-specific hypomethylated CpG loci.

(DOC)

References

- Matzuk MM, Lamb DJ (2008) The biology of infertility: research advances and clinical challenges. *Nat Med* 14: 1197–1213.
- World Health Organization (2010) WHO laboratory manual for the Examination and processing of human semen. Fifth Edition: WHO Press.
- Eckardstein SV, Schmidt A, Kamischke A, Simoni M, Gromoll J, et al. (2002) CAG repeat length in the androgen receptor gene and gonadotrophin suppression influence the effectiveness of hormonal male contraception. *Clin Endocrinol (Oxf)* 57: 647–655.
- Giachini C, Laface I, Guarducci E, Balercia G, Forti G, et al. (2008) Partial AZFc deletions and duplications: clinical correlates in the Italian population. *Hum Genet* 124: 399–410.
- Giachini C, Nuti F, Turner DJ, Laface I, Xue Y, et al. (2009) TSPY1 copy number variation influences spermatogenesis and shows differences among Y lineages. *J Clin Endocrinol Metab* 94: 4016–4022.
- Guarducci E, Nuti F, Becherini L, Rotondi M, Balercia G, et al. (2006) Estrogen receptor alpha promoter polymorphism: stronger estrogen action is coupled with lower sperm count. *Hum Reprod* 21: 994–1001.
- Huszar G, Vigue L, Corrales M (1988) Sperm creatine phosphokinase activity as a measure of sperm quality in normospermic, variablespermic, and oligospermic men. *Biol Reprod* 38: 1061–1066.
- Orlando C, Krausz C, Forti G, Casano R (1994) Simultaneous measurement of sperm LDH, LDH-X, CPK activities and ATP content in normospermic and oligozoospermic men. *Int J Androl* 17: 13–18.
- Gandini L, Lombardo F, Paoli D, Caruso F, Eleuteri P, et al. (2004) Full-term pregnancies achieved with ICSI despite high levels of sperm chromatin damage. *Hum Reprod* 19: 1409–1417.
- Jackson RE, Bormann CL, Hassun PA, Rocha AM, Motta EL, et al. (2010) Effects of semen storage and separation techniques on sperm DNA fragmentation. *Fertil Steril* 94: 2626–2630.
- Marchesi DE, Biederman H, Ferrera S, Herschlag A, Feng HL (2010) The effect of semen processing on sperm DNA integrity: comparison of two techniques using the novel Toluidine Blue Assay. *Eur J Obstet Gynecol Reprod Biol* 151: 176–180.
- Spano M, Cordelli E, Leter G, Lombardo F, Lenzi A, et al. (1999) Nuclear chromatin variations in human spermatozoa undergoing swim-up and cryopreservation evaluated by the flow cytometric sperm chromatin structure assay. *Mol Hum Reprod* 5: 29–37.
- Flanagan JM, Popendikyte V, Pozdniakovaite N, Sobolev M, Assadzadeh A, et al. (2006) Intra- and interindividual epigenetic variation in human germ cells. *Am J Hum Genet* 79: 67–84.
- Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, et al. (2009) Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 460: 473–478.
- Molaro A, Hodges E, Fang F, Song Q, McCombie WR, et al. (2011) Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 146: 1029–1041.
- Navarro-Costa P, Nogueira P, Carvalho M, Leal F, Cordeiro I, et al. (2010) Incorrect DNA methylation of the DAZL promoter CpG island associates with defective human sperm. *Hum Reprod* 25: 2647–2654.
- Boissonnas CC, Abdalaoui HE, Haelewyn V, Fauque P, Dupont JM, et al. (2010) Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* 18: 73–80.
- Hammoud SS, Nix DA, Hammoud AO, Gibson M, Cairns BR, et al. (2011) Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. *Hum Reprod* 26: 2558–2569.
- Hammoud SS, Purwar J, Pflueger C, Cairns BR, Carrell DT (2009) Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. *Fertil Steril* 94: 1728–1733.
- Houshdaran S, Cortessis VK, Siegmund K, Yang A, Laird PW, et al. (2007) Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PLoS One* 2: e1289.
- Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, et al. (2007) Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* 16: 2542–2551.
- Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, et al. (2008) Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* 14: 67–74.
- Marques CJ, Francisco T, Sousa S, Carvalho F, Barros A, et al. (2010) Methylation defects of imprinted genes in human testicular spermatozoa. *Fertil Steril* 94: 585–594.
- Pacheco SE, Houseman EA, Christensen BC, Marsit CJ, Kelsey KT, et al. (2011) Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS One* 6: e20280.
- Poplinski A, Tuttleman F, Kanber D, Horsthemke B, Gromoll J (2010) Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *Int J Androl* 33: 642–649.
- Aravin AA, Hannon GJ (2008) Small RNA silencing pathways in germ and stem cells. *Cold Spring Harb Symp Quant Biol* 73: 283–290.
- Deng W, Lin H (2002) miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* 2: 819–830.
- Kuramochi-Miyagawa S, Kimura T, Ijiri TW, Isobe T, Asada N, et al. (2004) Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* 131: 839–849.
- Juliano C, Wang J, Lin H (2011) Uniting Germline and Stem Cells: The Function of Piwi Proteins and the piRNA Pathway in Diverse Organisms. *Annu Rev Genet* 45: 447–469.
- Siddiqi S, Matushansky I (2011) Piwis and piwi-interacting RNAs in the epigenetics of cancer. *J Cell Biochem*.
- Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6: 692–702.
- Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, et al. (2011) High density DNA methylation array with single CpG site resolution. *Genomics* 98: 288–295.
- Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23: 257–258.
- Vavouri T, Lehner B Chromatin organization in sperm may be the major functional consequence of base composition variation in the human genome. *PLoS Genet* 7: e1002036.
- Carrell DT, Hammoud SS (2010) The human sperm epigenome and its potential role in embryonic development. *Mol Hum Reprod* 16: 37–47.
- Marques CJ, Carvalho F, Sousa M, Barros A (2004) Genomic imprinting in disruptive spermatogenesis. *Lancet* 363: 1700–1702.
- Brykczynska U, Hisano M, Erkek S, Ramos L, Oakeley EJ, et al. (2010) Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa. *Nat Struct Mol Biol* 17: 679–687.
- Krawetz SA, Kruger A, Lalancette C, Tagett R, Anton E, et al. (2011) A survey of small RNAs in human sperm. *Hum Reprod* 26: 3401–3412.
- Esposito T, Magliocca S, Formicola D, Gianfrancesco F (2011) piR_015520 belongs to Piwi-associated RNAs regulates expression of the human melatonin receptor 1A gene. *PLoS One* 6: e22727.

Table S6 Biological processes associated with genes linked to piRNAs specifically hypo or hypermethylated in spermatozoa.

(DOC)

Acknowledgments

We thank Paolo Sassone-Corsi (University of California, Irvine, USA), Jose Sanchez-Mut and Javier Carmona-Sanz (Idibell, Barcelona, Catalonia, Spain) for helpful discussions of the results. We thank Sue Hammoud and Doug Carrell (University of Utah, Salt lake City, USA) for providing us with their data regarding the histone-enriched loci. Sebastian Moran Salama y Carles Arribas Jorba (Idibell, Barcelona, Catalonia, Spain) are acknowledged for their technical assistance.

Author Contributions

Conceived and designed the experiments: CK ME. Performed the experiments: JS CC CG. Analyzed the data: CK JS SS. Contributed reagents/materials/analysis tools: CK ME HH. Wrote the paper: CK JS.

5. DISCUSSION

In the era of assisted reproductive technology (ART), previously infertile men can now generate their own progeny. Despite being a great advancement for the treatment of male infertility, the growing application of ART is indeed a double-edged sword because otherwise nontransmissible defects could be then transmitted to the future offspring. Though the prevalence of male infertility amounts to up to 7% of men of reproductive age, the etiology cannot be identified in up to half of cases. This deficit of understanding significantly limits the ability to counsel patients regarding prognosis for treatment; in particular, if not able to identify the cause for a man's infertility, it is impossible to tell patients the likelihood of infertility or other potential pathological conditions in ART- born offspring. Therefore, intense research has been and is still being done to detect novel factors involved in idiopathic male infertility. Considering the abundant number of genes involved in spermatogenesis (more than 2000) and that only a fraction of them has been analyzed so far, a genetic origin is highly probable. To date, few genetic defects have been found to have a clear-cut cause-effect relationship with male infertility, with the most important being Y-chromosome microdeletions and Klinefelter syndrome. The identification of additional genetic causes of male infertility will further improve the ability to appropriately diagnose and treat the disease. However, male infertility is a complex problem where not only genes, but also epigenetic factors seemingly play an important role. A number of studies focused on male infertility from the epigenetic standpoint, and there is now sufficient information supporting the idea that epigenetic changes do contribute to this condition. This thesis had the aim to investigate on the etiology of idiopathic male infertility by addressing both genetic and epigenetic aspects potentially involved. Two aspects were especially explored: the role of sex chromosomes-linked CNVs and the human sperm methylome.

5.1. X-linked CNVs

Copy number variations might lead to gene expression changes and thus result in phenotypically evident consequences (McCarroll et al., 2006; Nguyen et al., 2006; Repping et al., 2006), or they could just represent neutral variants responsible for the inter-individual variability. The phenotypic effect not always can be ascribed to the presence of genes within the region involved by the CNV (gene-dosage effect), but it can also be due to the alteration

of either the sequence or the position of non-coding genomic regions that regulate the expression of neighboring genes. The potential role of CNVs in complex diseases (Buchanan and Scherer, 2008) is fully supported by a growing number of data available in the literature (Wain et al., 2009; Choy et al., 2010; Fanciulli et al., 2010), which represent the fruit of the rapid development of ground-breaking technologies allowing the systematic analysis of the whole genome through microarrays (a-CGH, SNPs *arrays*) or high-throughput sequencing (*next generation sequencing*). Among the pathologies associated with CNVs are several neurological disturbances, autoimmune diseases, some types of cancer and even HIV-1 susceptibility. Diseases that develop as the consequence of an alteration in the genome causing the copy number reduction (or complete loss) or increase of dosage-sensitive genes, as well as the loss of their structural integrity, are referred to as “genomic disorders” (Lupski, 1998; Stankiewicz and Lupski, 2002); this term includes CNVs that involve genes and that have a clear clinical significance. The results of molecular studies on the model of genomic disorders provided evidence regarding the mechanism(s) for their recurrent origins. In contrast to allele-specific mutations, which generally originate from duplication errors or DNA mismatch repair, these rearrangements result from events of deletion, duplication, inversion and translocation and their generation is associated with mechanisms of recombination that mainly involve instable genomic architectures (Stankiewicz and Lupski, 2002), such as regions full of segmental duplications (SDs), favorable substrate for NAHR. In addition to SDs, repetitive sequences such as the retrotransposable L1 elements (LINE), *Alu* can act as NAHR substrates, if from similar families or with high enough sequence identity to facilitate homologous recombination. NAHR events occurring in the germline are responsible for more than 30 known genomic disorders (Sasaki et al., 2010), including Y-chromosome microdeletions, which are the second most frequent cause of male infertility. The research presented in this thesis in relation to the role of sex chromosomes-linked CNVs provides encouraging data regarding novel CNVs with potential implication in spermatogenic failure, thus representing an important step forward in the field of male infertility.

The analysis by a-CGH performed in our pilot study allowed us the identification of a consistent number of CNVs on the X chromosome, the majority of which (75.3%) were novel. Peculiarity of our study was the employment of a high-resolution (probe distance of 2-4 Kb) a-CGH platform specific for the X chromosome that allowed detecting even smaller CNVs (of 4-5 Kb size); another peculiar characteristic of our study was represented by the study population, in which only strictly selected subjects were included: 96 idiopathic infertile

patients with different grade of spermatogenic impairment (49 azoospermic, 25 cryptozoospermic and 22 oligozoospermic men) and 103 normozoospermic men.

By classifying CNVs in three size-based categories (<10 Kb, 10-100 Kb e >100 Kb) we showed that losses had the tendency of being smaller than gains: this difference became especially evident by assessing the prevalence of losses and gains belonging to the group of CNVs >100 Kb: it could be appreciated that 86% of CNVs was represented by gains, and the remaining 14% was represented by losses. This finding is not surprising if assuming that the loss of genetic material is probably more deleterious compared to a gain of a determined genomic fragment. From this standpoint, the tendency of deletions of showing reduced sizes can be indication of a greater pathogenicity of larger deletions, which presumably undergo negative selection and are thus less represented in the general population.

The main finding of this pilot study was the significantly higher burden of CNVs found in patients compared to controls, in terms of both mean number of CNVs/person (mainly dependent on an over-representation of losses) and mean size/person. Furthermore, a significantly lower sperm concentration and total sperm count was found in patients with >1 CNV compared to those with ≤ 1 CNV. This excess of X-linked CNVs and DNA loss in patients with reduced sperm count and the significant association between CNV number and sperm count in the infertile group supports a potential link between the observed CNV burden and spermatogenic failure. These conclusions are supported by two other genome-wide studies evaluating the involvement of CNVs in male infertility (Tüttelmann et al. 2011; Lopes et al. 2013). More specifically, Tüttelmann and colleagues (2011) reported a significant over-representation of sex-chromosomal CNVs in azoospermic men with SCO histology and a significant negative correlation between sperm count and the number of deletions at whole genome level, among normozoospermic men. More recently, Lopes and colleagues (2013) also reported a genome-wide CNV burden in azoospermic and oligozoospermic men replicating our finding of an X-linked CNV burden in men with spermatogenic failure (Krausz et al. 2012). From a clinical standpoint, of particular interest were patient-enriched (significantly more frequent in patients) and patient-specific (not found in controls) CNVs, since genes/regulatory elements within/nearby these regions presumably have a higher probability of being implicated in spermatogenic failure.

Following our a-CGH study, a number of patient-specific deletions and duplications were then selected for further investigation. As for deletions, the screening of a large group of patients and controls revealed that >90% of these deletions were private or rare (frequency <1%). Again, these findings are in line with previous whole genome array-CGH studies

(Tüttelmann et al., 2011; Stouffs et al., 2012); in the former, 27 patient-specific CNVs and only one recurrent CNV (gain) was found, whereas the latter reported that among 10 patient-specific autosomal CNVs, only two were recurrent. Furthermore, Lopes and colleagues (2013) reported the enrichment in large rare CNVs in men with spermatogenic failure. The role of rare CNVs has already been established for other multifactorial diseases (Manolio et al., 2009; Pinto et al., 2010) and since mutations causing spermatogenic disturbance are unlikely transmitted to the next generation, it is plausible that *de novo* rare mutations play a prominent role in primary testicular failure.

Of the abovementioned deletions, 3 recurrent deletions (frequency >1%) drew our attention for their exclusive (CNV67) or prevalent (CNV64 and CNV69) presence in patients. The analysis of these deletions was object of the second part of this thesis (Lo Giacco, Chianese et al. 2013.). A comprehensive investigation was performed including: i) the screening of a large series of strictly idiopathic patients (n=627) and normozoospermic controls (n=628) from two Mediterranean populations (Spanish and Italian); ii) the molecular characterization of deletions; iii) the exploration for functional elements in the region of interest; iv) a genotype-phenotype correlation analysis. All three deletions were mapping to the long arm of the X chromosome in q27.3 (CNV64) and q28 (CNV67 and CNV69). The smallest deletion was CNV64 removing at least 3.923 Kb, followed by CNV67, the minimum size of which was estimated to be 11.664 Kb, and CNV69 removing between 16.06-18.53 Kb. The alignment of the flanking sequences indicates that none of deletions originates from homologous recombination and that the most likely mechanism NHEJ.

Deletion carriers displayed a higher probability of having impaired spermatogenesis (OR=1.9 and 2.2 for CNV64 and CNV69, respectively) as well as sperm concentration/total motile sperm number was lower in carriers compared to non-carriers. These observations suggest that both CNV64 and CNV69 resulted significantly associated with spermatogenic failure. Interestingly, the molecular characterization of CNV69 deletion, revealed that at least two subtypes of CNV69 exist, named type A and type B. Since type B deletion was significantly more represented in patients than controls, there is a possibility that this deletion pattern may account for the potential deleterious effect of CNV69 on sperm production. This may be related to the closer position of type B deletion to an upstream insulator compared to type A. Importantly, through the breakpoint definition of type B deletion, we developed a simple PCR-based assay that will allow its screening to be effortlessly performed in other independent series of cases and controls. Despite the significant association observed between these deletions and the infertile phenotype, no genes were identified inside the

maximum size of the deletions. However, we also explored the regions surrounding each CNV for the presence of functional elements as, according to the ENCODE project, a very high percentage of non-coding DNA (80%) could have a biochemical function. A number of regulatory elements, including weak and strong enhancers, insulators and weak promoters were potentially affected because of their proximity to the deletion. For instance, large deletions may affect gene transcription also by changing 3D structure of chromatin leading to downstream effects on the regulation of protein coding regions.

The most interesting deletion was CNV67 because it was exclusively found in patients with a frequency of 1.1% ($p < 0.01$). The highly repetitive nature of the genomic region involved and the incomplete assembly of the currently available reference sequence of the human X chromosome prohibited a fine characterization of deletion breakpoints; notwithstanding, chromosome walking allowed a better definition of the deletion and lead to the hypothesis that the proximal copy of the *MAGE9A* gene -a CTA family member- and/or of its regulatory elements might be involved. Furthermore, considering that large-scale CNVs may also affect gene activity through a positional effect, the distal copy of *MAGE9A* may also be affected by CNV67. There is consistent evidence suggesting a potential involvement of CTA genes in spermatogenesis and of CTA gene dosage variation in spermatogenic impairment, therefore it is plausible that the loss of one *MAGE9A* copy would affect spermatogenesis explaining the phenotype observed in CNV67 carriers. Since CNV67 may also affect regulatory elements of another independently acquired X-linked multi-copy gene, *HSFX1/2* with testis-specific expression, the alteration of the expression of this gene may also account for the pathological semen phenotype observed in CNV67 carriers. Pedigree analyses of two CNV67 carriers indicated that CNV67 deletion is maternally inherited, thus not affecting female fertility in heterozygous state. This is in accordance with the lack of expression of both the *MAGE9A* and *HSFX1/2* in the ovary. Patient 11-041's family is especially informative since the pathological semen phenotype of the carrier (11-041) *versus* his normozoospermic non-carrier brother is a strong indicator for a pathogenic effect of the deletion on spermatogenesis.

The second line of investigation deriving from the initial pilot study focused on patient-specific duplications. Of 16 patient-specific duplications, five (DUP1A, DUP5, DUP20, DUP26 and DUP40) were selected for further investigation on a larger study population. Quantitative-PCR screening for these CNVs was performed in 276 idiopathic infertile patients and 327 controls in a conventional case-control setting (199 subjects belonged to the previous and a-CGH study). Our analysis revealed a significantly higher duplication load

in infertile patients compared to normozoospermic men. This data are in line with a recent study by Tüttelmann et al (2012), who provided evidence for the first statistically significant duplication burden on the X chromosome and reported a significantly higher number of gains in azoospermic patients compared to both oligozoospermic patients and normal controls.

While four of the five CNVs (DUP5, DUP20, DUP26 and DUP40) did not individually reach statistical significance, they remained patient-specific. All CNVs include, or are in close proximity to, genes with testis-specific expression and potential implication in spermatogenesis. One gain (DUP1A) was found at a significantly higher frequency in patients. This gain contains a long non-coding RNA (*LINC00685*) that potentially acts as a negative regulator of a gene with potential role in spermatogenesis, *PPP2R3B*. This proposed mechanism could not be confirmed by functional studies because *in vitro* human spermatogenic cell culture is not available and the *LINC00685* antisense is not present in easily accessible model organisms such as mouse and *Drosophila*. However, expression data provides evidence for an inverse relationship between *PPP2R3B* and *LINC00685* levels in a number of different tissues (Chalmel et al., 2007, 2012). Concerning the testis, samples with a high expression of *PPP2R3B* show comparatively low expression of *LINC00685* (Chalmel 2007, 2012). Based on this observation, we propose that the mechanism by which DUP1A could lead to spermatogenic failure is through increased negative regulation, caused by the duplicated *LINC00685* that would decrease *PPP2R3B* transcription in the developing germ cells. Although the role of *PPP2R3B* in spermatogenesis has not been explored, indirect functional evidence supports its involvement both in mitosis and meiosis. To further support our hypothesis, we reanalyzed raw data deposited in dbVar and found that in the study by Tüttelmann et al. a number of CNVs (5 gains and 2 losses) were found in the PAR1 of men suffering azoospermia due to SCOS. A second mechanism by which large PAR1-linked CNVs, like DUP1A and DUP5, may lead to impaired spermatogenesis is represented by a structural effect disturbing male meiosis i.e. altering the recombination event that occurs between the PAR1 regions of the sex chromosomes (Chandley, 1989; Mohandas et al., 1992). PAR1 recombination becomes progressively more frequent towards the distal telomeric boundary (Filatov, 2004), where DUP1A and DUP5 are located, showing the importance of this region during meiosis.

5.2. Y-chromosome microdeletions and *SHOX* CNVs.

Y-chromosome microdeletions are the most frequent genetic cause of male infertility, second only to the Klinefelter Syndrome. Hence, the molecular diagnosis of these deletions has become an important test in the diagnostic workup of the infertile male (Vogt et al., 1996; Krausz and Degl'Innocenti, 2006; Krausz et al., 2014). For some couples with male partners carrying Y-chromosome microdeletions, ART represents the only opportunity to father their own biological children. However, genetic defects on the Y chromosome will be inevitably transmitted to a male progeny; therefore, current EAA/EMQN guidelines assess that genetic counseling is mandatory before the couple undergoes ICSI/IVF treatment. But, what should these couples be counseled for?

Until now, the answer to this question would have been one and certain: since deletions occurring in the father's Y chromosome will be unavoidably transmitted to the son, the latter will suffer from impaired sperm production. Of numerous articles focused on Y-chromosome microdeletions, only a few aimed to define whether Y chromosome deletions might lead to other pathological conditions, beside spermatogenic failure. Concerns have been raised about the potential risk for Turner's syndrome (45,X) in the offspring and other phenotypic anomalies associated with 45X/46XY mosaic karyotype, including ambiguous genitalia (Papadimas et al., 2001; Patsalis et al., 2002, 2005; Papanikolaou et al., 2003; Tian et al., 2012). However, the number of reported ICSI babies born from fathers affected by Yq microdeletions is roughly 50, thus still relatively low (Kleiman et al.; Kent-First et al., 1996; Mulhall et al., 1997; Jiang et al., 1999; Kamischke et al., 1999; van Golde et al., 2001; Oates et al., 2002; Peterlin et al., 2002; Choi et al., 2004; Cram et al., 2004; Stouffs et al., 2005; Kihaila et al., 2005; Simoni et al., 2008; Mau Kai et al., 2008; Mateu et al., 2010; Lo Giacco et al., 2013a). From these data, it can be evinced that apparently children are phenotypically normal - except for one case of pulmonary atresia and a hypoplastic right ventricle (Page et al., 1999) - and no ambiguous genitalia or cases of Turner syndrome have been observed among them.

To sound a note of warning in this regard was the paper by Jorgez et al. 2011 reporting that Y-chromosome microdeletions might not only cause spermatogenic failure but also increase the risk for PAR-related pathologies, especially emphasizing *SHOX* gene involvement. The *SHOX* gene is the best-established disease locus in PAR1 and aberrations in this gene account for 3.2% of patients with isolated short stature, 89% of patients with Léri- Weill-

dyschondrosteosis and 100% of patients with Langer-mesomelic-dysplasia (Belin et al., 1998). The severest consequences are due to *SHOX* haploinsufficiency, which can cause short stature as well as Turner skeletal abnormalities such as short fourth metacarpals, *cubitus valgus* and characteristics of Léri-Weill dyschondrosteosis (LWD; OMIM #127300) (Kosho et al 1999). If Y-chromosome microdeletions carriers truly were at risk of incurring also *SHOX*-linked pathologies, genetic counseling would need to change radically; for instance, infertile couples undergoing ART would need to be informed that their sons will have not only fertility problems but also a higher risk of developing *SHOX*-related disorders. This would then require the compulsory testing for *SHOX*- linked CNVs in men carrying Yq microdeletions.

The most alarming finding reported by Jorgez et al. was that on a total of 87 men with Yq microdeletions, five samples had *SHOX* haploinsufficiency: four carrying a normal karyotype and one carrying an idic(Y); however, the authors did not specify whether these patients also had any phenotypic features typical of *SHOX* haploinsufficiency. Given the relevance and the potential clinical impact of this issue, we performed a multicenter investigation on a large study population in order to investigate whether the hypothesis of an association between Yq microdeletions and *SHOX* haploinsufficiency could be confirmed. Our study provides the largest collection of men carrying Y-chromosome microdeletions analyzed so far in association with *SHOX* copy number. The inclusion of partial Y-chromosome rearrangements, such as gr/gr deletions and gr/gr deletion-duplications, is an important novelty of our study, as it is relevant the inclusion of a very large number of men with complete AZFc microdeletions (4-fold higher than the previous study), who represent the more plausible candidates for ART attempts. Our data seemingly indicate that *SHOX* CNVs are mainly linked to the individual's karyotype rather than the mere presence of microdeletions. For instance, among men with normal karyotype we found that only 1.1% (2/177) Y-chromosome deletion carriers displayed duplications of the *SHOX* gene; however, apart from a relatively high stature (though still within the normal range) the men's clinical work-up revealed no other pathological conditions than azoospermia. Our study had the limitation that the number of controls analyzed was rather low to assess whether these duplications represented a neutral polymorphism or are actually associated with the presence of the microdeletion.

The most important and reassuring finding of our study was that none of the men with Yq microdeletions had *SHOX* deletions, utterly contrasting the aforementioned hypothesis that microdeletion carriers are at a higher risk of developing pathologies caused by *SHOX* haploinsufficiency. The reasons underlying the discrepancy between our study and the

previous publication might be related to either methodological or ethnical issues. However, a proper comparison was made difficult by the succinct description of the methods employed in the previous study and of the ethnic background of their study population.

5.3. The “normal” sperm methylome

The influence of the paternal epigenome on embryo development has long been sidelined. However, abundant scientific evidence has been produced to support that the peculiar nature of the sperm epigenetic landscape might play a larger role in development than previously thought. Among the diverse epigenetic modifications, DNA methylation represents an important signaling tool involved in mammalian development. After fertilization, the genomes inherited from both spermatozoa and oocytes undergo a massive wave of nearly complete demethylation, and then lineage-specific patterns of *de novo* methylation occur during or after gastrulation. Information nowadays available suggests that this epigenetic reprogramming is associated with re-establishment of the gamete developmental potential, correct initiation of embryonic gene expression and early lineage development in the embryo. The understanding of this complex process and the description of the sperm DNA methylome have both evolutionary and clinical (Molaro et al., 2011; Carrell, 2012) implications. The first description of the entire sperm DNA methylome was based on the analysis of whole semen samples of only two sperm donors. Studies including a larger study population do exist, but were based either on the analysis of few genes or on low-resolution arrays (Kobayashi et al., 2007; Marques et al., 2008, 2010; Hammoud et al., 2009, 2010; Navarro-Costa et al., 2010b; Poplinski et al., 2010; Pacheco et al., 2011).

In this regard, our study represents an important step forward. For instance, in order to obtain a comprehensive overview of the sperm DNA methylome, we used a 487,317 CpG-feature microarray - the larger performed so far - to investigate on the DNA methylation profile in eight normozoospermic men.

The first question we wanted to address was whether quality-fractionated sperm populations coming from the same subject had different methylation patterns. With the advent of ART, sperm selection methods have been developed to obtain an enrichment of a subpopulation of spermatozoa exhibiting the best structural and functional characteristics, indicative of optimal fertilizing ability. Whether quality differences between sperm subpopulations were mirroring differences in the DNA methylation pattern was still an unexplored field. Below expectation, we observed a strong stability of the methylation patterns between sperm fractions enriched in “poor” and “good” quality spermatozoa. Any difference was observed

neither in genes for which DNA methylation defects had been previously reported in association with impaired sperm production/quality, including 45 imprinted genes, already available on the array, and 10 genes additionally selected from the literature. These data indicate that in normozoospermic men, structural and functional differences between sperm subpopulations do not derive from DNA methylation defects. It is plausible that environmental factors altering the histone codes during adult spermatogenesis are responsible for the anomalies observed in the “poor quality” sperm fraction. We also hypothesize that previous observations on the association between infertility and abnormal sperm methylation profile reflect a defect that occurs during fetal life or in early puberty and thus similarly to the normozoospermic data, will be present in all sperm quality fractions of a given infertile man.

The significant epigenetic variability previously observed in human germ cells (Flanagan et al., 2006) led us to the second question addressed in our study: is there a difference between DNA methylation profiles of different normozoospermic subjects’ sperm subpopulations? In contrast with the previous observations, our data clearly proved that the methylation pattern in different individuals with similar sperm characteristics was highly conserved. The meticulous attention given to the sperm selection process, by which we obtained high-purity sperm fractions, gave security that our data did not result from the presence of contaminating somatic cells. The high correlation found in the selected fraction enriched with *good*-quality spermatozoa might suggest that spermatozoa destined to fertilize show a highly stable sperm methylation profile between normozoospermic individuals, irrespective of differences in life style, age and BMI.

In order to provide a general description and comparative analyses of the sperm DNA methylome, we focused on the fraction enriched with *good*-quality spermatozoa showing a complete lack of significant inter-individual differences. Interestingly, the “normal” sperm DNA methylome appeared highly polarized towards the two extreme of DNA methylation, with 96% of CpGs being either hypomethylated (20%) or hypermethylated (80%). Hypo- and hypermethylated loci were highly conserved among different individuals reaching to a concordance of 95% for hypomethylated CpGs and to 83.3% for the hypermethylated ones. Of note, we observed a high inter-individual conservation of hypomethylated loci, especially abundant in promoter regions (63.6%), suggesting that normal spermatogenesis requires strictly controlled methylation levels in specific gene promoters. Among conserved hypomethylated loci, an exceptionally high number was linked to the X and Y chromosomes, supporting the importance of sex chromosome-linked genes in the spermatogenic process.

Another notable finding was that conserved hypomethylated CpGs mapping inside histone-enriched regions were associated with genes involved in developmental processes. Accordingly, the majority of developmental gene promoters available in the array were mapping inside histone - retained regions. Our data are outright consistent with what reported in the literature. It is known that the histone-protamine replacement occurs in a incomplete manner, with about 5–15% of chromatin remaining bound to nucleosomes (Tanphaichitr et al., 1978; Wykes and Krawetz, 2003). Interestingly, this phenomenon was found to be programmatic and not simply a causality, suggesting that retained histones might play a role in epigenetic regulation (Arpanahi et al., 2009; Hammoud et al., 2009). In proven fertile patients, histone retention is found at the promoters of genes important in the embryo including developmental gene promoters, microRNA clusters, and imprinted loci, suggesting the programmatic nature of nucleosome retention is programmatic in nature (Hammoud et al. 2009).

Our comprehensive description of the sperm DNA methylome included the comparison of the latter with the methylation profile of somatic cells. Differential methylation could be found only in 4.6% of CpGs and was mainly due to the overrepresentation of hypomethylated loci in spermatozoa. The analysis of functional genomic distribution, CpG content and associated RNA transcripts in differentially methylated sites revealed that promoters of genes related to sperm-specific hypomethylated CpGs were strongly associated with gamete generation. Further, hypomethylated genes mapping to histone-retained regions were found to be involved in developmental processes and spermatogenesis, further highlighting that sperm histone retention is naturally programmed to rapidly activate genes involved in embryonic development.

6. CONCLUSIONS

Aim 1:

- A significantly higher X chromosome-linked CNV burden was observed in idiopathic infertile patients and its association with lower sperm counts indicates a potential link to spermatogenic failure.
- The X chromosome-linked recurrent deletions CNV64, CNV67, CNV69 are significantly associated with spermatogenic failure. CNV67 specificity to impaired spermatogenesis and its frequency of 1.1% in oligo/azoospermic men make this deletion particularly interesting, since it resembles the AZF regions on the Y chromosome with potential clinical implications.
- The X chromosome-linked recurrent gain DUP1a is significantly associated with spermatogenic failure and two possible mechanisms have been provided to explain the pathogenesis of the associated infertile phenotype: one relating a gene dosage effect due to *LINC00685/PPP2R3B* misbalance, and the other due to a potential structural effect in the PAR1.

Aim 2:

- Both partial and complete microdeletions in men with normal karyotype are unlikely associated with *SHOX* haploinsufficiency, reassuring that the only established risk for ART offspring born from men with Yq microdeletions remains spermatogenic impairment.

Aim 3:

- The DNA methylation profile is highly conserved among normozoospermic subjects, indicating a lack of inter-individual variability.
- Different quality-fractioned sperm subpopulations deriving from the same individual have a stable methylation profile. This data suggest that the acquisition of the sperm methylation status is a process that occurs during early development of the testis, and thus *poor* quality spermatozoa, present in a normozoospermic ejaculate, is not related to altered methylation status.
- The majority of CpGs are either hypomethylated or hypermethylated, indicating a stable polarization of the sperm methylation profile.

- Hypomethylated or hypermethylated loci show a distinct genomic and functional organization.
- Histone-retained hypomethylated loci are associated with embryonic development, further supporting the importance of sperm epigenetic regulation in the mammalian development.

7. BIBLIOGRAPHY

- Adamson KA, Cross I, Batch JA, Rappold GA, Glass IA, Ball SG (2002) Trisomy of the short stature homeobox-containing gene (SHOX), resulting from a duplication-deletion of the X chromosome. *Clin Endocrinol (Oxf)* 56:671–675.
- Aitman TJ et al. (2006) Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature* 439:851–855.
- Almeida LG, Sakabe NJ, deOliveira AR, Silva MCC, Mundstein AS, Cohen T, Chen Y-T, Chua R, Gurung S, Gnjatic S, Jungbluth AA, Caballero OL, Bairoch A, Kiesler E, White SL, Simpson AJG, Old LJ, Camargo AA, Vasconcelos ATR (2009) CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res* 37:D816–D819.
- Arpanahi A, Brinkworth M, Iles D, Krawetz SA, Paradowska A, Platts AE, Saida M, Steger K, Tedder P, Miller D (2009) Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Res* 19:1338–1349.
- Aston KI, Carrell DT (2009) Genome-wide study of single-nucleotide polymorphisms associated with azoospermia and severe oligozoospermia. *J Androl* 30:711–725.
- Aston KI, Krausz C, Laface I, Ruiz-Castané E, Carrell DT (2010) Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent. *Hum Reprod* 25:1383–1397.
- Aston KI, Punj V, Liu L, Carrell DT (2012) Genome-wide sperm deoxyribonucleic acid methylation is altered in some men with abnormal chromatin packaging or poor in vitro fertilization embryogenesis. *Fertil Steril* 97:285–292.
- Bacolla A, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeysinghe SS, O’Connell CD, Cooper DN, Wells RD (2004) Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci U S A* 101:14162–14167.
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte R V, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. *Science* 297:1003–1007.
- Bassaganyas L, Riveira-Muñoz E, García-Aragonés M, González JR, Cáceres M, Armengol L, Estivill X (2013) Worldwide population distribution of the common LCE3C-LCE3B deletion associated with psoriasis and other autoimmune disorders. *BMC Genomics* 14:261.
- Belin V, Cusin V, Viot G, Girlich D, Toutain A, Moncla A, Vekemans M, Le Merrer M, Munnich A, Cormier-Daire V (1998) SHOX mutations in dyschondrosteosis (Leri-Weill syndrome). *Nat Genet* 19:67–69.

- Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, Brown LG, Rozen S, Warren WC, Wilson RK, Page DC (2010) Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* 466:612–616.
- Benchaib M, Braun V, Ressenkoff D, Lornage J, Durand P, Niveleau A, Guérin JF (2005) Influence of global sperm DNA methylation on IVF results. *Hum Reprod* 20:768–773.
- Benito-Sanz S, Royo JL, Barroso E, Paumard-Hernández B, Barreda-Bonis AC, Liu P, Gracia R, Lupski JR, Campos-Barros Á, Gómez-Skarmeta JL, Heath KE (2012) Identification of the first recurrent PAR1 deletion in Léri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. *J Med Genet* 49:442–450.
- Boissonnas CC, Abdalaoui H El, Haelewyn V, Fauque P, Dupont JM, Gut I, Vaiman D, Jouannet P, Tost J, Jammes H (2010) Specific epigenetic alterations of IGF2-H19 locus in spermatozoa from infertile men. *Eur J Hum Genet* 18:73–80.
- Bosch E, Jobling MA (2003) Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Hum Mol Genet* 12:341–347.
- Brandell RA, Mielnik A, Liotta D, Ye Z, Veeck LL, Palermo GD, Schlegel PN (1998) AZFb deletions predict the absence of spermatozoa with testicular sperm extraction: preliminary report of a prognostic genetic test. *Hum Reprod* 13:2812–2815.
- Bronner C, Chataigneau T, Schini-Kerth VB, Landry Y (2007) The “Epigenetic Code Replication Machinery”, ECREM: a promising drugable target of the epigenetic cell memory. *Curr Med Chem* 14:2629–2641.
- Bruder CEG et al. (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* 82:763–771.
- Buchanan JA, Scherer SW (2008) Contemplating effects of genomic structural variation. *Genet Med* 10:639–647.
- Burgoyne PS, Evans EP, Holland K (1983) XO monosomy is associated with reduced birthweight and lowered weight gain in the mouse. *J Reprod Fertil* 68:381–385.
- Cairns J, Foster PL (1991) Adaptive reversion of a frameshift mutation in *Escherichia coli*. *Genetics* 128:695–701.
- Carrell DT (2012) Epigenetics of the male gamete. *Fertil Steril* 97:267–274.
- Chalmel F, Lardenois A, Evrard B, Mathieu R, Feig C, Demougin P, Gattiker A, Schulze W, Jégou B, Kirchhoff C, Primig M (2012) Global human tissue profiling and protein network analysis reveals distinct levels of transcriptional germline-specificity and identifies target genes for male infertility. *Hum Reprod* 27:3233–3248.
- Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SSW, Demougin P, Gattiker A, Moore J, Patard J-J, Wolgemuth DJ, Jégou B, Primig M (2007) The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A* 104:8346–8351.

- Chandley AC (1989) Asymmetry in chromosome pairing: a major factor in de novo mutation and the production of genetic disease in man. *J Med Genet* 26:546–552.
- Chianese C, Gunning AC, Giachini C, Daguin F, Balercia G, Ars E, Lo Giacco D, Ruiz-Castañé E, Forti G, Krausz C (2014) X chromosome-linked CNVs in male infertility: discovery of overall duplication load and recurrent, patient-specific gains with potential clinical relevance. *PLoS One* 9:e97746.
- Choi JM, Chung P, Veeck L, Mielnik A, Palermo GD, Schlegel PN (2004) AZF microdeletions of the Y chromosome and in vitro fertilization outcome.
- Choy KW, Setlur SR, Lee C, Lau TK (2010) The impact of human copy number variation on a new era of genetic testing. *BJOG* 117:391–398.
- Clement-Jones M, Schiller S, Rao E, Blaschke RJ, Zuniga A, Zeller R, Robson SC, Binder G, Glass I, Strachan T, Lindsay S, Rappold GA (2000) The short stature homeobox gene SHOX is involved in skeletal abnormalities in Turner syndrome. *Hum Mol Genet* 9:695–702.
- Cook EH, Scherer SW (2008) Copy-number variations associated with neuropsychiatric conditions. *Nature* 455:919–923.
- Cram D, Lynch M, O'Bryan MK, Salvado C, McLachlan RI, de Kretser DM (2004) Genetic screening of infertile men. *Reprod Fertil Dev* 16:573–580.
- Cruger DG, Agerholm I, Byriel L, Fedder J, Bruun-Petersen G (2003) Genetic analysis of males from intracytoplasmic sperm injection couples. *Clin Genet* 64:198–203.
- D'haene B, Hellemans J, Craen M, De Schepper J, Devriendt K, Fryns J-P, Keymolen K, Debals E, de Klein A, de Jong EM, Segers K, De Paepe A, Mortier G, Vandesompele J, De Baere E (2010) Improved molecular diagnostics of idiopathic short stature and allied disorders: quantitative polymerase chain reaction-based copy number profiling of SHOX and pseudoautosomal region 1. *J Clin Endocrinol Metab* 95:3010–3018.
- Darai-Ramqvist E, Sandlund A, Müller S, Klein G, Imreh S, Kost-Alimova M (2008) Segmental duplications and evolutionary plasticity at tumor chromosome break-prone regions. *Genome Res* 18:370–379.
- Doerksen T (1996) Developmental exposure of male germ cells to 5-azacytidine results in abnormal preimplantation development in rats. *Biol Reprod* 55:1155–1162.
- Duan Z, Duan Y, Lamendola DE, Yusuf RZ, Naeem R, Penson RT, Seiden M V (2003) Overexpression of MAGE/GAGE genes in paclitaxel/doxorubicin-resistant human cancer cell lines. *Clin Cancer Res* 9:2778–2785.
- Eichler EE (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet* 17:661–669.
- Ellison JW, Wardak Z, Young MF, Gehron Robey P, Laig-Webster M, Chiong W (1997) PHOG, a candidate gene for involvement in the short stature of Turner syndrome. *Hum Mol Genet* 6:1341–1347.

- Fanciulli M et al. (2007) FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet* 39:721–723.
- Fanciulli M, Petretto E, Aitman TJ (2010) Gene copy number variation and common human disease. *Clin Genet* 77:201–213.
- Ferguson-Smith MA, Lennox B, Mack WS, Stewart JS (1957) Klinefelter's syndrome; frequency and testicular morphology in relation to nuclear sex. *Lancet* 273:167–169.
- Filatov DA (2004) A gradient of silent substitution rate in the human pseudoautosomal region. *Mol Biol Evol* 21:410–417.
- Flanagan JM, Pependikyte V, Pozdniakovaite N, Sobolev M, Assadzadeh A, Schumacher A, Zangeneh M, Lau L, Virtanen C, Wang S-C, Petronis A (2006) Intra- and interindividual epigenetic variation in human germ cells. *Am J Hum Genet* 79:67–84.
- Fratta E, Coral S, Covre A, Parisi G, Colizzi F, Danielli R, Nicolay HJM, Sigalotti L, Maio M (2011) The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Mol Oncol* 5:164–182.
- Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ (2004) Complex SNP-related sequence variation in segmental genome duplications. *Nat Genet* 36:861–866.
- Fridlyand J, Snijders AM, Ylstra B, Li H, Olshen A, Segraves R, Dairkee S, Tokuyasu T, Ljung BM, Jain AN, McLennan J, Ziegler J, Chin K, Devries S, Feiler H, Gray JW, Waldman F, Pinkel D, Albertson DG (2006) Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 6:96.
- Gervasini C, Grati FR, Lalatta F, Tabano S, Gentilin B, Colapietro P, De Toffol S, Frontino G, Motta F, Maitz S, Bernardini L, Dallapiccola B, Fedele L, Larizza L, Miozzo M (2010) SHOX duplications found in some cases with type I Mayer-Rokitansky-Kuster-Hauser syndrome. *Genet Med* 12:634–640.
- Glynn SA, Gammell P, Heenan M, O'Connor R, Liang Y, Keenan J, Clynes M (2004) A new superinvasive in vitro phenotype induced by selection of human breast carcinoma cells with the chemotherapeutic drugs paclitaxel and doxorubicin. *Br J Cancer* 91:1800–1807.
- Grigeliuniene G, Schoumans J, Neumeyer L, Ivarsson A, Eklöf O, Enkvist O, Tordai P, Fosdal I, Myhre AG, Westphal O, Nilsson NO, Elfving M, Ellis I, Anderlid BM, Fransson I, Tapia-Paez I, Nordenskjöld M, Hagenäs L, Dumanski JP (2001) Analysis of short stature homeobox-containing gene (SHOX) and auxological phenotype in dyschondrosteosis and isolated Madelung deformity. *Hum Genet* 109:551–558.
- Gu W, Zhang F, Lupski JR (2008) Mechanisms for human genomic rearrangements. *Pathogenetics* 1:4.
- Hammer M, Zegura S (2002) The human Y chromosome haplogroup tree: nomenclature and phylogeny of its major division. *Annu Rev Anthr* 31:303–321.

- Hammoud SS, Nix DA, Zhang H, Purwar J, Carrell DT, Cairns BR (2009) Distinctive chromatin in human sperm packages genes for embryo development. *Nature* 460:473–478.
- Hammoud SS, Purwar J, Pflueger C, Cairns BR, Carrell DT (2010) Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility. *Fertil Steril* 94:1728–1733.
- Handy DE, Castro R, Loscalzo J (2011) Epigenetic modifications: basic mechanisms and role in cardiovascular disease. *Circulation* 123:2145–2156.
- Hansson K, Szuhai K, Knijnenburg J, van Haeringen A, de Pater J (2007) Interstitial deletion of 6q without phenotypic effect. *Am J Med Genet A* 143A:1354–1357.
- Helena Mangs A, Morris BJ (2007) The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Curr Genomics* 8:129–136.
- Herceg Z, Vaissière T (2011) Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics* 6:804–819.
- Hochstenbach R, Hackstein JH (2000) The comparative genetics of human spermatogenesis: clues from flies and other model organisms. *Results Probl Cell Differ* 28:271–298.
- Houshdaran S, Cortessis VK, Siegmund K, Yang A, Laird PW, Sokol RZ (2007) Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm. *PLoS One* 2:e1289.
- Hu Z et al. (2012) A genome-wide association study in Chinese men identifies three risk loci for non-obstructive azoospermia. *Nat Genet* 44:183–186.
- Hurles ME, Dermitzakis ET, Tyler-Smith C (2008) The functional impact of structural variation in humans. *Trends Genet* 24:238–245.
- Hurst LD (2001) Evolutionary genomics. Sex and the X. *Nature* 411:149–150.
- Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36:949–951.
- Jackson JA, Fink GR (1981) Gene conversion between duplicated genetic elements in yeast. *Nature* 292:306–311.
- Jakobsson M et al. (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451:998–1003.
- Jiang MC, Lien YR, Chen SU, Ko TM, Ho HN, Yang YS (1999) Transmission of de novo mutations of the deleted in azoospermia genes from a severely oligozoospermic male to a son via intracytoplasmic sperm injection. *Fertil Steril* 71:1029–1032.
- Jobling MA (2008) Copy number variation on the human Y chromosome. *Cytogenet Genome Res* 123:253–262.

- Jorgez CJ, Weedin JW, Sahin A, Tannour-Louet M, Han S, Bournat JC, Mielnik A, Cheung SW, Nangia AK, Schlegel PN, Lipshultz LI, Lamb DJ (2011) Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. *J Clin Endocrinol Metab* 96:E674–E679.
- Kamischke A, Gromoll J, Simoni M, Behre HM, Nieschlag E (1999) Transmission of a Y chromosomal deletion involving the deleted in azoospermia (DAZ) and chromodomain (CDY1) genes from father to son through intracytoplasmic sperm injection: case report. *Hum Reprod* 14:2320–2322.
- Kamp C, Huellen K, Fernandes S, Sousa M, Schlegel PN, Mielnik A, Kleiman S, Yavetz H, Krause W, Küpker W, Johannisson R, Schulze W, Weidner W, Barros A, Vogt PH (2001) High deletion frequency of the complete AZFa sequence in men with Sertoli-cell-only syndrome. *Mol Hum Reprod* 7:987–994.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830–838.
- Kelly TLJ, Li E, Trasler JM 5-aza-2'-deoxycytidine induces alterations in murine spermatogenesis and pregnancy outcome. *J Androl* 24:822–830.
- Kent-First MG, Kol S, Muallem A, Ofir R, Manor D, Blazer S, First N, Itskovitz-Eldor J (1996) The incidence and possible relevance of Y-linked microdeletions in babies born after intracytoplasmic sperm injection and their infertile fathers. *Mol Hum Reprod* 2:943–950.
- Khazamipour N, Noruzinia M, Fatehmanesh P, Keyhanee M, Pujol P (2009) MTHFR promoter hypermethylation in testicular biopsies of patients with non-obstructive azoospermia: the role of epigenetics in male infertility. *Hum Reprod* 24:2361–2364.
- Kihaile PE, Yasui A, Shuto Y (2005) Prospective assessment of Y-chromosome microdeletions and reproductive outcomes among infertile couples of Japanese and African origin. *J Exp Clin Assist Reprod* 2:9.
- Kim PM, Lam HYK, Urban AE, Korbel JO, Affourtit J, Grubert F, Chen X, Weissman S, Snyder M, Gerstein MB (2008) Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res* 18:1865–1874.
- Kleiman SE, Almog R, Yogev L, Hauser R, Lehavi O, Paz G, Yavetz H, Botchan A (2012) Screening for partial AZFa microdeletions in the Y chromosome of infertile men: is it of clinical relevance? *Fertil Steril* 98:43–47.
- Kleiman SE, Yogev L, Gamzu R, Hauser R, Botchan A, Paz G, Lessing JB, Yaron Y, Yavetz H Three-generation evaluation of Y-chromosome microdeletion. *J Androl* 20:394–398.
- Kobayashi H, Sato A, Otsu E, Hiura H, Tomatsu C, Utsunomiya T, Sasaki H, Yaegashi N, Arima T (2007) Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients. *Hum Mol Genet* 16:2542–2551.

- Kosho T, Muroya K, Nagai T, Fujimoto M, Yokoya S, Sakamoto H, Hirano T, Terasaki H, Ohashi H, Nishimura G, Sato S, Matsuo N, Ogata T (1999) Skeletal features and growth patterns in 14 patients with haploinsufficiency of SHOX: implications for the development of Turner syndrome. *J Clin Endocrinol Metab* 84:4613–4621.
- Kosova G, Scott NM, Niederberger C, Prins GS, Ober C (2012) Genome-wide association study identifies candidate genes for male fertility traits in humans. *Am J Hum Genet* 90:950–961.
- Krausz C (2011) Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 25:271–285.
- Krausz C, Degl'Innocenti S (2006) Y chromosome and male infertility: update, 2006. *Front Biosci* 11:3049–3061.
- Krausz C, Forti G (2000) Clinical aspects of male infertility. *Results Probl Cell Differ* 28:1–21.
- Krausz C, Giachini C, Lo Giacco D, Daguin F, Chianese C, Ars E, Ruiz-Castane E, Forti G, Rossi E (2012a) High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PLoS One* 7:e44887.
- Krausz C, Hoefsloot L, Simoni M, Tüttelmann F (2014) EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: state-of-the-art 2013. *Andrology* 2:5–19.
- Krausz C, Quintana-Murci L, Barbaux S, Siffroi JP, Rouba H, Delafontaine D, Souleyreau-Therville N, Arvis G, Antoine JM, Erdei E, Taar JP, Tar A, Jeandidier E, Plessis G, Bourgeron T, Dadoune JP, Fellous M, McElreavey K (1999) A high frequency of Y chromosome deletions in males with nonidiopathic infertility. *J Clin Endocrinol Metab* 84:3606–3612.
- Krausz C, Quintana-Murci L, McElreavey K (2000) Prognostic value of Y deletion analysis: what is the clinical prognostic value of Y chromosome microdeletion analysis? *Hum Reprod* 15:1431–1434.
- Krausz C, Sandoval J, Sayols S, Chianese C, Giachini C, Heyn H, Esteller M (2012b) Novel insights into DNA methylation features in spermatozoa: stability and peculiarities. *PLoS One* 7:e44479.
- Kurahashi H, Emanuel BS (2001) Long AT-rich palindromes and the constitutional t(11;22) breakpoint. *Hum Mol Genet* 10:2605–2617.
- Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC (2001) The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* 29:279–286.
- Lam K-WG, Jeffreys AJ (2006) Processes of copy-number change in human DNA: the dynamics of {alpha}-globin gene deletion. *Proc Natl Acad Sci U S A* 103:8921–8927.

- Lam K-WG, Jeffreys AJ (2007) Processes of de novo duplication of human alpha-globin genes. *Proc Natl Acad Sci U S A* 104:10950–10955.
- Larsen F, Gundersen G, Prydz H (1992) Choice of enzymes for mapping based on CpG islands in the human genome. *Genet Anal Tech Appl* 9:80–85.
- Lee C, Scherer SW (2010) The clinical context of copy number variation in the human genome. *Expert Rev Mol Med* 12:e8.
- Lee JA, Carvalho CMB, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 131:1235–1247.
- Lercher MJ, Urrutia AO, Hurst LD (2003) Evidence that the human X chromosome is enriched for male-specific but not female-specific genes. *Mol Biol Evol* 20:1113–1116.
- Levy S et al. (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5:e254.
- Lieber MR (2008) The mechanism of human nonhomologous DNA end joining. *J Biol Chem* 283:1–5.
- Liu L, Rando TA (2011) Manifestations and mechanisms of stem cell aging. *J Cell Biol* 193:257–266.
- Liu Y, Lu C, Yang Y, Fan Y, Yang R, Liu C-F, Korolev N, Nordenskiöld L (2011) Influence of histone tails and H4 tail acetylations on nucleosome-nucleosome interactions. *J Mol Biol* 414:749–764.
- Lo Giacco D, Chianese C, Sánchez-Curbelo J, Bassas L, Ruiz P, Rajmil O, Sarquella J, Vives A, Ruiz-Castañé E, Oliva R, Ars E, Krausz C (2013a) Clinical relevance of Y-linked CNV screening in male infertility: new insights based on the 8-year experience of a diagnostic genetic laboratory. *Eur J Hum Genet* Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24193344> [Accessed January 8, 2014].
- Lo Giacco D, Chianese C, Sanchez-Curbelo, Josvany Bassas L, Ruiz P, Osvaldo R, Sarquella J, Vives A, Ruiz-Castañé, Eduard Oliva R, Ars E, Krausz C (2013b) Clinical relevance of Y-linked CNV screening in male infertility: new insights based on the 8-year experience of a diagnostic genetic laboratory. *Eur J Hum Genet*.
- Lopes AM et al. (2013) Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and Both Sex Chromosomes, including the Gene DMRT1. *PLoS Genet* 9:e1003349.
- Lupski JR (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* 14:417–422.
- Lupski JR, Stankiewicz P (2005) Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet* 1:e49.
- Manolio TA et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753

- Marchini A, Marttila T, Winter A, Caldeira S, Malanchi I, Blaschke RJ, Häcker B, Rao E, Karperien M, Wit JM, Richter W, Tommasino M, Rappold GA (2004) The short stature homeodomain protein SHOX induces cellular growth arrest and apoptosis and is expressed in human growth plate chondrocytes. *J Biol Chem* 279:37103–37114
- Marques CJ, Carvalho F, Sousa M, Barros A (2004) Genomic imprinting in disruptive spermatogenesis. *Lancet* 363:1700–1702.
- Marques CJ, Costa P, Vaz B, Carvalho F, Fernandes S, Barros A, Sousa M (2008) Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia. *Mol Hum Reprod* 14:67–74.
- Marques CJ, Francisco T, Sousa S, Carvalho F, Barros A, Sousa M (2010) Methylation defects of imprinted genes in human testicular spermatozoa. *Fertil Steril* 94:585–594.
- Mateu E, Rodrigo L, Martínez MC, Peinado V, Milán M, Gil-Salom M, Martínez-Jabaloyas JM, Remohí J, Pellicer A, Rubio C (2010) Aneuploidies in embryos and spermatozoa from patients with Y chromosome microdeletions. *Fertil Steril* 94:2874–2877.
- Matsuda Y, Moens PB, Chapman VM (1992) Deficiency of X and Y chromosomal pairing at meiotic prophase in spermatocytes of sterile interspecific hybrids between laboratory mice (*Mus domesticus*) and *Mus spretus*. *Chromosoma* 101:483–492.
- Matzuk MM, Lamb DJ (2002) Genetic dissection of mammalian fertility pathways. *Nat Cell Biol* 4 Suppl:s41–s49.
- Mau Kai C, Juul A, McElreavey K, Ottesen AM, Garn ID, Main KM, Loft A, Jørgensen N, Skakkebaek NE, Andersen AN, Rajpert-De Meyts E (2008) Sons conceived by assisted reproduction techniques inherit deletions in the azoospermia factor (AZF) region of the Y chromosome and the DAZ gene copy number. *Hum Reprod* 23:1669–1678.
- McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM (2006) Common deletion polymorphisms in the human genome. *Nat Genet* 38:86–92.
- McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24:529–538.
- Mohandas TK, Speed RM, Passage MB, Yen PH, Chandley AC, Shapiro LJ (1992) Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. *Am J Hum Genet* 51:526–533.
- Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD (2011) Sperm Methylation Profiles Reveal Features of Epigenetic Inheritance and Evolution in Primates. *Cell* 146:1029–1041.
- Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, David C, Turner JMA (2009) genes exhibiting post-meiotic expression. 40:794–799.

- Mueller JL, Mahadevaiah SK, Park PJ, Warburton PE, Page DC, Turner JMA (2008) The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet* 40:794–799.
- Mueller JL, Skaletsky H, Brown LG, Zaghul S, Rock S, Graves T, Auger K, Warren WC, Wilson RK, Page DC (2013) Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* 45:1083–1087.
- Mulhall JP, Reijo R, Alagappan R, Brown L, Page D, Carson R, Oates RD (1997) Azoospermic men with deletion of the DAZ gene cluster are capable of completing spermatogenesis: fertilization, normal embryonic development and pregnancy occur when retrieved testicular spermatozoa are used for intracytoplasmic sperm injection. *Hum Reprod* 12:503–508.
- Mumm S, Molini B, Terrell J, Srivastava A, Schlessinger D (1997) Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 60-kb resolution. *Genome Res* 7:307–314.
- Munns CJF, Haase HR, Crowther LM, Hayes MT, Blaschke R, Rappold G, Glass IA, Batch JA (2004) Expression of SHOX in human fetal and childhood growth plate. *J Clin Endocrinol Metab* 89:4130–4135.
- Nanassy L, Carrell DT (2011) Abnormal methylation of the promoter of CREM is broadly associated with male factor infertility and poor sperm quality but is improved in sperm selected by density gradient centrifugation. *Fertil Steril* 95:2310–2314.
- Navarro-Costa P, Gonçalves J, Plancha CE (2010a) The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility. *Hum Reprod Update* 16:525–542.
- Navarro-Costa P, Nogueira P, Carvalho M, Leal F, Cordeiro I, Calhaz-Jorge C, Gonçalves J, Plancha CE (2010b) Incorrect DNA methylation of the DAZL promoter CpG island associates with defective human sperm. *Hum Reprod* 25:2647–2654.
- Navarro-Costa P, Plancha CE, Gonçalves J (2010c) Genetic dissection of the AZF regions of the human Y chromosome: thriller or filler for male (in)fertility? *J Biomed Biotechnol*:936569.
- Ng MH, Wong IH, Lo KW (1999) DNA methylation changes and multiple myeloma. *Leuk Lymphoma* 34:463–472.
- Nguyen D-Q, Webber C, Ponting CP (2006) Bias of selection on human copy-number variants. *PLoS Genet* 2:e20.
- Nuti F, Krausz C (2008) Gene polymorphisms/mutations relevant to abnormal spermatogenesis. *Reprod Biomed Online* 16:504–513.
- Oates RD, Silber S, Brown LG, Page DC (2002) Clinical characterization of 42 oligospermic or azoospermic men with microdeletion of the AZFc region of the Y chromosome, and of 18 children conceived via ICSI. *Hum Reprod* 17:2813–2824.

- Odorisio T, Rodriguez TA, Evans EP, Clarke AR, Burgoyne PS (1998) The meiotic checkpoint monitoring synapsis eliminates spermatocytes via p53-independent apoptosis. *Nat Genet* 18:257–261.
- Ogata T, Matsuo N, Nishimura G (2001) SHOX haploinsufficiency and overdosage: impact of gonadal function status. *J Med Genet* 38:1–6.
- Ohno S (1967) Sex chromosomes and sex-linked genes. (Monographs on endocrinology, Vol. 1.).
- Pacheco SE, Houseman EA, Christensen BC, Marsit CJ, Kelsey KT, Sigman M, Boekelheide K (2011) Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS One* 6:e20280.
- Page DC (1987) Hypothesis: a Y-chromosomal gene causes gonadoblastoma in dysgenetic gonads. *Development* 101 Suppl:151–155.
- Page DC, Silber S, Brown LG (1999) Men with infertility caused by AZFc deletion can produce sons by intracytoplasmic sperm injection, but are likely to transmit the deletion and infertility. *Hum Reprod* 14:1722–1726.
- Papadimas J, Goulis DG, Giannouli C, Papanicolaou A, Tarlatzis B, Bontis JN (2001) Ambiguous genitalia, 45,X/46,XY mosaic karyotype, and Y chromosome microdeletions in a 17-year-old man. *Fertil Steril* 76:1261–1263.
- Papanikolaou AD, Goulis DG, Giannouli C, Gounioti C, Bontis JN, Papadimas J (2003) Intratubular germ cell neoplasia in a man with ambiguous genitalia, 45,X/46,XY mosaic karyotype, and Y chromosome microdeletions. *Endocr Pathol* 14:177–182.
- Park J-H, Lee S-W (2002) Hypertonicity induction of melanoma antigen, a tumor-associated antigen. *Mol Cells* 13:288–295.
- Patsalis PC, Sismani C, Quintana-Murci L, Taleb-Bekkouché F, Krausz C, McElreavey K (2002) Effects of transmission of Y chromosome AZFc deletions. *Lancet* 360:1222–1224
- Patsalis PC, Skordis N, Sismani C, Kousoulidou L, Koumbaris G, Eftychi C, Stavrides G, Ioulianos A, Kitsiou-Tzeli S, Galla-Voumvouraki A, Kosmaïdou Z, Hadjiathanasiou CG, McElreavey K (2005) Identification of high frequency of Y chromosome deletions in patients with sex chromosome mosaicism and correlation with the clinical phenotype and Y-chromosome instability. *Am J Med Genet A* 135:145–149.
- Pawelczak KS, Turchi JJ (2008) A mechanism for DNA-PK activation requiring unique contributions from each strand of a DNA terminus and implications for microhomology-mediated nonhomologous DNA end joining. *Nucleic Acids Res* 36:4022–4031.
- Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, Park HS, Kim J-I, Seo J-S, Yakhini Z, Laderman S, Bruhn L, Lee C (2008) The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* 82:685–695.

- Peterlin B, Kunej T, Sinkovec J, Gligorievska N, Zorn B (2002) Screening for Y chromosome microdeletions in 226 Slovenian subfertile men. *Hum Reprod* 17:17–24.
- Pinto D et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466:368–372.
- Piotrowski A, Bruder CEG, Andersson R, Diaz de Ståhl T, Menzel U, Sandgren J, Poplawski A, von Tell D, Crasto C, Bogdan A, Bartoszewski R, Bebok Z, Krzyzanowski M, Jankowski Z, Partridge EC, Komorowski J, Dumanski JP (2008) Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat* 29:1118–1124.
- Poplinski A, Tüttelmann F, Kanber D, Horsthemke B, Gromoll J (2010) Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1. *Int J Androl* 33:642–649.
- Portela A, Esteller M (2010) Epigenetic modifications and human disease. *Nat Biotechnol* 28:1057–1068.
- Raicu F, Popa L, Apostol P, Cimponeriu D, Dan L, Ilinca E, Dracea LL, Marinescu B, Gavrila L (2003) Screening for microdeletions in human Y chromosome--AZF candidate genes and male infertility. *J Cell Mol Med* 7:43–48.
- Rajpert-De Meyts E, Ottesen AM, Garn ID, Aksglaede L, Juul A (2011) Deletions of the Y chromosome are associated with sex chromosome aneuploidy but not with Klinefelter syndrome. *Acta Paediatr* 100:900–902.
- Rao E, Weiss B, Fukami M, Rump A, Niesler B, Mertz A, Muroya K, Binder G, Kirsch S, Winkelmann M, Nordsiek G, Heinrich U, Breuning MH, Ranke MB, Rosenthal A, Ogata T, Rappold GA (1997) Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat Genet* 16:54–63.
- Redon R et al. (2006) Global variation in copy number in the human genome. *Nature* 444:444–454.
- Repping S, van Daalen SKM, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC, Rozen S (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet* 38:463–467.
- Rodríguez-Santiago B, Brunet A, Sobrino B, Serra-Juhé C, Flores R, Armengol L, Vilella E, Gabau E, Guitart M, Guillamat R, Martorell L, Valero J, Gutiérrez-Zotes A, Labad A, Carracedo A, Estivill X, Pérez-Jurado LA (2010) Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia. *Mol Psychiatry* 15:1023–1033.
- Roos L, Brøndum Nielsen K, Tümer Z (2009) A duplication encompassing the SHOX gene and the downstream evolutionarily conserved sequences. *Am J Med Genet A* 149A:2900–2901.
- Rosenfeld RG (2001) A SHOX to the system. *J Clin Endocrinol Metab* 86:5672–5673.

- Ross MT et al. (2005) The DNA sequence of the human X chromosome. *Nature* 434:325–337.
- Sasaki M, Lange J, Keeney S (2010) Genome destabilization by homologous recombination in the germ line. *Nat Rev Mol Cell Biol* 11:182–195.
- Sato A, Hiura H, Okae H, Miyauchi N, Abe Y, Utsunomiya T, Yaegashi N, Arima T (2011) Assessing loss of imprint methylation in sperm from subfertile men using novel methylation polymerase chain reaction Luminex analysis. *Fertil Steril* 95:129–134, 134.e1–e4.
- Saus E, Brunet A, Armengol L, Alonso P, Crespo JM, Fernández-Aranda F, Guitart M, Martín-Santos R, Menchón JM, Navinés R, Soria V, Torrens M, Urretavizcaya M, Vallès V, Gratacòs M, Estivill X (2010) Comprehensive copy number variant (CNV) analysis of neuronal pathways genes in psychiatric disorders identifies rare variants within patients. *J Psychiatr Res* 44:971–978.
- Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Distèche C, McGillivray B, de la Chapelle A, Page DC (1998) Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum Mol Genet* 7:1–11.
- Sebat J et al. (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525–528.
- Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff V V, Albertson DG, Pinkel D, Eichler EE (2005) Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 77:78–88.
- Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet* 13 Spec No:R57–R64.
- Siffroi JP, Le Bourhis C, Krausz C, Barbaux S, Quintana-Murci L, Kanafani S, Rouba H, Bujan L, Bourrouillou G, Seifer I, Boucher D, Fellous M, McElreavey K, Dadoune JP (2000) Sex chromosome mosaicism in males carrying Y chromosome long arm deletions. *Hum Reprod* 15:2559–2562.
- Simoni M, Tüttelmann F, Gromoll J, Nieschlag E (2008) Clinical consequences of microdeletions of the Y chromosome: the extended Münster experience. *Reprod Biomed Online* 16:289–303.
- Skaletsky H et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.
- Slack A, Thornton PC, Magner DB, Rosenberg SM, Hastings PJ (2006) On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genet* 2:e48.
- Song R, Ro S, Michaels JD, Park C, McCarrey JR, Yan W (2009) Many X-linked microRNAs escape meiotic sex chromosome inactivation. *Nat Genet* 41:488–493.

- Song X, Zhao Y, Cai Q, Zhang Y, Niu Y (2013) Association of the Glutathione S-transferases M1 and T1 polymorphism with male infertility: a meta-analysis. *J Assist Reprod Genet* 30:131–141.
- Stankiewicz P, Lupski JR (2002) Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genet Dev* 12:312–319.
- Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. *Annu Rev Med* 61:437–455.
- Stouffs K, Lissens W (2012) X chromosomal mutations and spermatogenic failure. *Biochim Biophys Acta* 1822:1864–1872.
- Stouffs K, Lissens W, Tournaye H, Haentjens P (2011) What about gr/gr deletions and male infertility? Systematic review and meta-analysis. *Hum Reprod Update* 17:197–209.
- Stouffs K, Lissens W, Tournaye H, Van Steirteghem A, Liebaers I (2005) The choice and outcome of the fertility treatment of 38 couples in whom the male partner has a Yq microdeletion. *Hum Reprod* 20:1887–1896.
- Stouffs K, Tournaye H, Liebaers I, Lissens W (2009) Male infertility and the involvement of the X chromosome. *Hum Reprod Update* 15:623–637.
- Stouffs K, Vandermaelen D, Massart A, Menten B, Vergult S, Tournaye H, Lissens W (2012) Array comparative genomic hybridization in male infertility. *Hum Reprod* 27:921–929.
- Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* 33:25–35.
- Tachdjian G, Aboura A, Portnoï M-F, Pasquier M, Bourcigaux N, Simon T, Rousseau G, Finkel L, Benkhalifa M, Christin-Maitre S (2008) Cryptic Xp duplication including the SHOX gene in a woman with 46,X, del(X)(q21.31) and premature ovarian failure. *Hum Reprod* 23:222–226.
- Tanphaichitr N, Sobhon P, Taluppeth N, Chalermisarachai P (1978) Basic nuclear proteins in testicular cells and ejaculated spermatozoa in man. *Exp Cell Res* 117:347–356.
- Thomas NS, Harvey JF, Bunyan DJ, Rankin J, Grigeliuniene G, Bruno DL, Tan TY, Tomkins S, Hastings R (2009) Clinical and molecular characterization of duplications encompassing the human SHOX gene reveal a variable effect on stature. *Am J Med Genet A* 149A:1407–1414.
- Tian L, Zhang J-W, Shen C-X, Du Y, Zhou X (2012) [Cytogenetics and Y chromosome AZF microdeletions in infertile patients with mosaic karyotype Klinefelter syndrome (46,XY/47,XXY/48, XXYY/49,XXXXY)]. *Zhonghua Nan Ke Xue* 18:545–550.
- Tiepolo L, Zuffardi O (1976) Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet* 34:119–124.

- Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, Beck S, Hurles ME (2008) Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet* 40:90–95.
- Tüttelmann F, Laan M, Grigorova M, Punab M, Söber S, Gromoll J (2012) Combined effects of the variants FSHB -211G>T and FSHR 2039A>G on male reproductive parameters. *J Clin Endocrinol Metab* 97:3639–3647.
- Tüttelmann F, Rajpert-De Meyts E, Nieschlag E, Simoni M (2007) Gene polymorphisms and male infertility--a meta-analysis and literature review. *Reprod Biomed Online* 15:643–658.
- Tüttelmann F, Simoni M, Kliesch S, Ledig S, Dworniczak B, Wieacker P, Röpke A (2011) Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PLoS One* 6:e19426.
- Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson M V, Eichler EE (2005) Fine-scale structural variation of the human genome. *Nat Genet* 37:727–732.
- Tyler-Smith C, Oakey RJ, Larin Z, Fisher RB, Crocker M, Affara NA, Ferguson-Smith MA, Muenke M, Zuffardi O, Jobling MA (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nat Genet* 5:368–375.
- Ungerer M, Knezovich J, Ramsay M (2013) In utero alcohol exposure, epigenetic changes, and their consequences. *Alcohol Res* 35:37–46.
- Van Golde RJ, Wetzels AM, de Graaf R, Tuerlings JH, Braat DD, Kremer JA (2001) Decreased fertilization rate and embryo quality after ICSI in oligozoospermic men with microdeletions in the azoospermia factor c region of the Y chromosome. *Hum Reprod* 16:289–292.
- Visser L, Westerveld GH, Korver CM, van Daalen SKM, Hovingh SE, Rozen S, van der Veen F, Repping S (2009) Y chromosome gr/gr deletions are a risk factor for low semen quality. *Hum Reprod* 24:2667–2673.
- Vogt PH et al. (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Mol Genet* 5:933–943
- Wain L V, Armour JAL, Tobin MD (2009) Genomic copy number variation, human health, and disease. *Lancet* 374:340–350.
- Wang PJ, McCarrey JR, Yang F, Page DC (2001) An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 27:422–426.
- Wang PJ, Page DC, McCarrey JR (2005) Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. *Hum Mol Genet* 14:2911–2918.

- Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39:457–466.
- Wei B, Xu Z, Ruan J, Zhu M, Jin K, Zhou D, Xu Z, Hu Q, Wang Q, Wang Z (2012) MTHFR 677C>T and 1298A>C polymorphisms and male infertility risk: a meta-analysis. *Mol Biol Rep* 39:1997–2002.
- Weterings E, Chen DJ (2008) The endless tale of non-homologous end-joining. *Cell Res* 18:114–124.
- Weterings E, van Gent DC (2004) The mechanism of non-homologous end-joining: a synopsis of synopsis. *DNA Repair (Amst)* 3:1425–1435.
- WHO (2010) World Health Organization Laboratory Manual for the Examination and Processing of Human Semen.
- Willcocks LC, Lyons PA, Clatworthy MR, Robinson JI, Yang W, Newland SA, Plagnol V, McGovern NN, Condliffe AM, Chilvers ER, Adu D, Jolly EC, Watts R, Lau YL, Morgan AW, Nash G, Smith KGC (2008) Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J Exp Med* 205:1573–1582.
- Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL (2007) A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* 80:91–104.
- Wu W, Shen O, Qin Y, Niu X, Lu C, Xia Y, Song L, Wang S, Wang X (2010) Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolate reductase (MTHFR). *PLoS One* 5:e13884.
- Wykes SM, Krawetz SA (2003) The structural organization of sperm chromatin. *J Biol Chem* 278:29471–29477.
- Zhao H et al. (2012) A genome-wide association study reveals that variants within the HLA region are associated with risk for nonobstructive azoospermia. *Am J Hum Genet* 90:900–906.
- Zheng K, Yang F, Wang PJ (2010) Regulation of male fertility by X-linked genes. *J Androl* 31:79–85.
- Zhu J, Yao X (2009) Use of DNA methylation for cancer detection: promises and challenges. *Int J Biochem Cell Biol* 41:147–154.

Annexes

Krausz C and **Chianese C**. Genetic testing and counselling for male infertility. *Curr Opin Endocrinol Diabetes Obes*. 2014 Jun;21(3):244-50.

Krausz C, **Chianese C**, Giachini C, Guarducci E, Laface I, Forti G. The Y chromosome-linked copy number variations and male fertility. *J Endocrinol Invest*. 2011 May;34(5):376-82.

Chiara Chianese, Sara Brilli and Csilla Krausz. Genomic Changes in Spermatozoa of the Aging Male. *Adv Exp Med Biol*. 2014. 2014;791:13-26.



Genetic testing and counselling for male infertility

Csilla Krausz and Chiara Chianese

Purpose of review

Genetic disorders can be identified in about 15% of cases of male infertility. With the widespread application of assisted reproductive technology, infertile patients are now given the possibility of having their biological children; however, a genetic risk exists for assisted reproductive technology-born offspring, implying the necessity for future parents to be appropriately informed about potential consequences. In this review, we provide current recommendations on clinical genetic testing and genetic counselling.

Recent findings

New insights are presented concerning Klinefelter syndrome, X and Y chromosome-linked deletions, monogenic diseases and pharmacogenetics.

Summary

As for Klinefelter patients, novel preventive measures to preserve fertility have been proposed although they are not yet applicable in the routine setting. Y-chromosome deletions have both diagnostic and prognostic values and their testing is advised to be performed according to the new European Academy of Andrology/European Molecular Genetics Quality Network guidelines. Among monogenic diseases, major advances have been obtained in the identification of novel genes of hypogonadotrophic hypogonadism. Pharmacogenetic approaches of hormonal treatment in infertile men with normal values of follicle-stimulating hormone (FSH) are promising and based on *FSHR* and *FSHB* polymorphisms. X chromosome-linked deletions are relevant for impaired spermatogenesis. In about 40% of male infertility, the cause is unknown and novel genetic factors are expected to be discovered in the near future.

Keywords

azoospermia, genetics, Klinefelter syndrome, male infertility, Y chromosome

INTRODUCTION

Nearly 15% of couples wishing to conceive seek medical treatment for infertility. The male factor accounts for approximately half of these involuntarily childless partners and in at least 15% of cases it is related to genetic disorders, including both chromosomal and single-gene alterations [1]. Genetic causes can be detected in all major etiologic categories of male infertility (pretesticular, testicular and post-testicular forms) and genetic tests became part of the routine diagnostic workup in selected groups of patients. With the widespread application of assisted reproductive technology (ART), infertile men are now given the possibility of having their biological children; however, a genetic risk exists for ART-born offspring, implying the necessity for future parents to be appropriately informed on potential consequences. In this review, we provide an outline of the most recent innovative aspects of the genetics of male infertility with a primary focus on clinical genetic testing (Table 1) and genetic counselling of the infertile couple.

CHROMOSOMAL ANOMALIES

Chromosomal abnormalities can be either numerical or structural, and the anomalies most frequently found in relationship with male infertility are Klinefelter syndrome and translocations (Robertsonian and reciprocal) or inversions. A general consensus exists on the correlation between the severity of the testicular phenotype and a higher frequency of chromosomal anomalies, thus karyotype analysis is recommended in Europe for men with fewer than 10 million spermatozoa per milliliter [2]. This cutoff for testing has been

Department of Experimental and Clinical Biomedical Sciences, University of Florence, Viale Pieraccini, Florence, Italy

Correspondence to Csilla Krausz, Department of Experimental and Clinical Biomedical Sciences, University of Florence, Viale Pieraccini, 6, Florence, 50139, Italy. Tel: +39 055 4271487; fax: +39 055 4271371; e-mail: c.krausz@dfc.unifi.it

Curr Opin Endocrinol Diabetes Obes 2014, 21:244–250

DOI:10.1097/MED.0000000000000058

KEY POINTS

- Genetic factors can be identified in about 15% of infertile men and, because approximately 2000 genes are implicated in male fertility, idiopathic cases are likely related to genetic/epigenetic factors acting either alone or in combination with environmental factors.
- Testicular tissue and SSCs banking have been proposed as an option for Klinefelter syndrome, the most frequent chromosomal anomaly in azoospermic men, but cannot be recommended on a routine basis.
- Kallmann syndrome and normosmic congenital hypogonadotrophic hypogonadism are currently considered different phenotypic expression of the same disease, which is apparently inherited in an oligogenic rather than a Mendelian fashion and characterized by the occurrence of reversibility in a subset of patients.
- The definition of the extension of the AZFa and AZFb deletions is strongly recommended by the European Academy of Andrology/European Molecular Genetics Quality Network guidelines because of its prognostic relevance for sperm retrieval in patients undergoing assisted reproductive technology.
- *FSHB* and *FSHR* polymorphisms are important determinants of male fitness and their analysis opens new perspectives toward a pharmacogenetic approach in FSH therapy in oligozoospermic men with normal to low serum FSH level, although it cannot be recommended yet in a routine clinical practice.

established on the observations that patients with fewer than 10 million spermatozoa per milliliter show a 10-fold higher incidence (4%) of mainly autosomal structural abnormalities compared with the general population. The highest frequency of karyotype anomalies has been observed in nonobstructive azoospermic men (15–16%) and Klinefelter syndrome accounts alone for 14% of cases, whereas the remaining are structural anomalies [2]. In 80–90% of Klinefelter patients, a 47,XXY karyotype is encountered, whereas the remaining cases can display a mosaic karyotype (46,XY/47,XXY), additional X chromosomes (e.g., 48,XXXY) or structurally abnormal X chromosomes. Klinefelter men usually suffer from azoospermia and thus are unable to spontaneously conceive. Nowadays, testicular sperm retrieval from microdissection testicular sperm extraction (TESE) may detect residual foci of active spermatogenesis in the testes of azoospermic 47,XXY adult men [3]. Sperm retrieved by microdissection TESE can be used for subsequent intracytoplasmic sperm injection (ICSI) procedure. However, the positive

outcome of ART in Klinefelter patients is threatened by the fact that the potential of successful sperm retrieval decreases with age and after testosterone therapies [4]. Consequently, new approaches of fertility preservation have been introduced in the counselling of pubertal Klinefelter patients. Because germ cell depletion starts with the onset of puberty, spermatogonial stem cell (SSC) banking at early puberty has been proposed as a strategy to preserve the fertility of these patients. Two recent studies focused on this matter: a histological study investigating whether cryopreservation of SSCs might benefit to Klinefelter boys [5] and the other proposing a combined clinical-hormonal strategy to detect early spermatogenesis for further SSCs retrieval from a single testicular biopsy [6]. Data by van Saen *et al.* [5] suggest that for optimal SSCs preservation, spermatogonia should be retrieved preferably before testis hyalinization occurs. However, their results show that it is rather difficult to find spermatogonia in tubules with a normal architecture even in pubertal boys, as massive fibrosis and hyalinization was observed in all, except one, patients over a total of seven patients. Likely, the hyalinization process might progress very rapidly in some Klinefelter adolescents; therefore, the basis for a maximum SSCs preservation would be the early detection of the syndrome. Furthermore, Gies *et al.* [6], investigating on seven Klinefelter adolescents, failed in the attempt to determine the optimal timing for SSCs retrieval, because neither clinical nor hormonal parameters were sufficiently predictive. A third study [4] claims the possibility of sperm retrieval for fertility preservation from semen samples of Klinefelter adolescents ($n=8$) before the administration of hormone replacement treatments, and suggest bilateral testicular biopsy for TESE or tissue freezing in case of TESE failure only if azoospermia was confirmed after two/three semen collection attempts. However, only in one boy (with a mosaic karyotype) spermatozoa could be recovered in the ejaculate. Of the five patients undergoing TESE, mature spermatozoa could be retrieved only in one (16 years old), whereas in another (15.5 years old) only immature sperm cells could be recovered.

Overall, data available in the literature seem to not fully support the feasibility of fertility preservation in Klinefelter adolescents. It is worth noting, though, that the number of patients analyzed overall is rather small ($n = 22$), because the diagnosis of Klinefelter is rarely made prepubertally. Still, the retrieval rate of mature germ cells seems rather low and SSCs cryopreservation would appear as the only approach; however, germ cells *in vitro* maturation techniques are still at an experimental

Table 1. Outline of the genetic testing currently included in the diagnostic workup of male infertility

Phenotypic category	Indication for testing	Genetic target	Test
Hypogonadotrophic hypogonadism	Kallmann syndrome and normosmic hypogonadotrophic hypogonadism	<i>KAL1, FGFR1, FGFR3, PROKR2, PROKR2, CHD7, KISS1/KISS1R, TAC3/TACR3, GNRH1/GNRHR, NELF, FGFR1, IL17RD, SOX10, DUSP6, SPRY4, FLRT3, SEMA3A</i>	Mutational screening through direct sequencing
	Isolated gonadotrophin deficiency	<i>FSH and LH</i>	
Primary testicular dysfunction	Azoospermia or <10 millions sp./ml	Chromosomal anomalies	Detection of numerical and structural alterations through karyotype analysis
	Azoospermia or <5 millions sp./ml	Y-chromosome AZF regions	Detection of AZF microdeletions (AZFa, AZFb, AZFc) through a PCR \pm method using the recommended set of STSs (EAA/EMQN guidelines [1])
	Oligozoospermia (<20 millions sp./ml)	AZFc region ^a	Detection of gr/gr deletion through a combined method including the analysis of recommended STSs and <i>DAZ/CDY1</i> gene dosage (EAA/EMQN guidelines [1])
	Hypoandrogenization due to PAIS	<i>AR</i>	Screening for <i>AR</i> mutations (http://androgendb.mcgill.ca/)
Congenital obstruction	Oligo/azoospermia due to CAVD	<i>CFTR</i>	Mutational screening through a mutation panel including the more frequent <i>CFTR</i> mutations specific to a given population; second step analysis through direct sequencing

CAVD, Congenital Absence of the Vas Deferens (uni/bilateral): All men with CAVD should be screened for *CFTR* mutation except for patients with renal agenesis/malformation; PAIS, Partial Androgen Insensitivity Syndrome; sp., spermatozoa.

^aRecommended in countries where solid data exist in terms of risk for oligozoospermia.

stage in animal models and far from guaranteeing future fertility in humans. Further studies are needed in order to clarify whether SSCs banking could be considered a valid approach to the fertility preservation of Klinefelter patients.

Genetic counselling

Structural anomalies might become unbalanced in the offspring [7] with serious health consequences and for this reason preimplantation genetic diagnosis (PGD) or prenatal diagnosis is highly advisable as preventive measures concerning Klinefelter syndrome, although studies of sperm fluorescence in-situ hybridization (FISH) in these patients revealed an increased rate of both gonosomal and autosomal (especially chromosomes 13, 18 and 21) aneuploidy in their spermatozoa, overall data are quite reassuring that children born from Klinefelter patients apparently do not demonstrate a higher risk for aneuploidies [8]. Notwithstanding, couples are generally offered PGD in order to ensure that the embryo selected for transfer is karyotypically normal.

Y-CHROMOSOME MICRODELETIONS

AZF deletions on the Yq represent the most frequent molecular genetic cause of impaired spermatogenesis. There are five recurrent deletion types (generally called AZFa, AZFb, AZFb+c and AZFc) and their frequency varies according to the semen phenotype, reaching the highest frequency in idiopathic azoospermia (approximately 10%) (Fig. 1a). Indications for routine testing include men with fewer than 5 million spermatozoa per milliliter although the large majority of carriers have fewer than 2 million spermatozoa per milliliter [10^a,11]. In the novel European Academy of Andrology/European Molecular Genetics Quality Network guidelines for Y-chromosome microdeletions [10^a], it is now clearly stated that the definition of the extension of AZFa and AZFb deletions with the so-called 'second-step' markers is of clinical relevance, because carriers of large deletions have virtually zero chance of successful sperm retrieval in the testis. The role of the gr/gr deletion, removing half of the gene content of the AZFc region, has been widely discussed during the last 10 years. Thanks to the publication of four meta-analyses, there is now a general acceptance

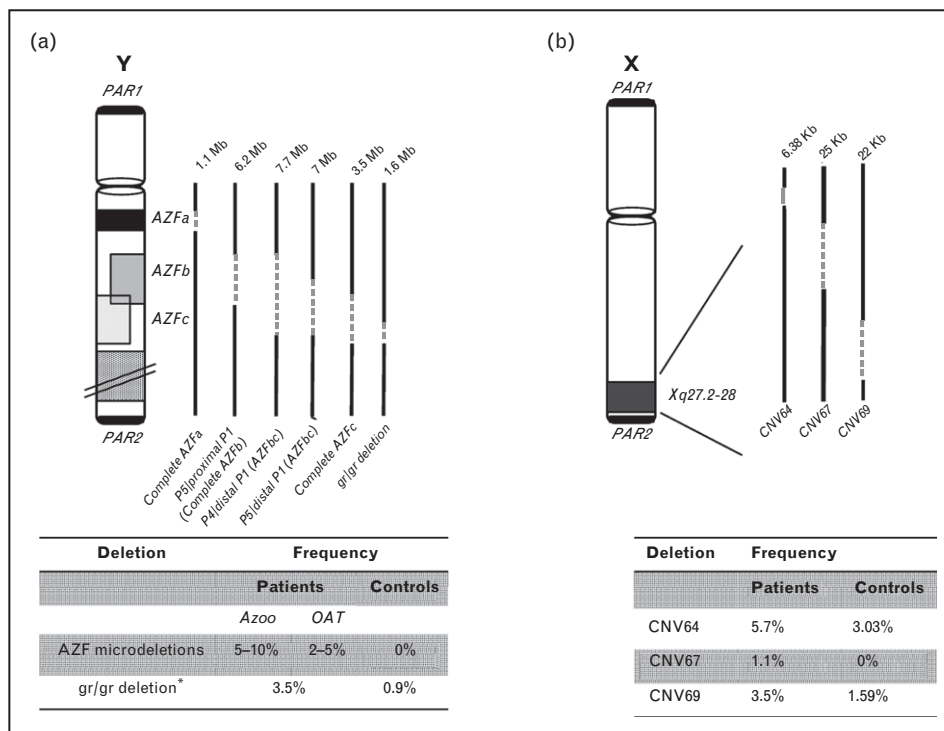


FIGURE 1. Deletions on the sex chromosomes. (a) Schematic representation of the Y chromosome showing the localization of the three azoospermia factor regions and the two pseudoautosomal regions (PAR1 and PAR2). On the right side, a model of the six recurrent deletions occurring on the Yq is provided. The thinner lines indicate the deleted portions. In the upper part, the genomic size (Mb) of each deletion is specified. In the lower part, AZF microdeletions and gr/gr deletion frequencies in patients and controls are reported. Azoo, azoospermic; OAT, oligoasthenoteratozoospermic. *Mean frequencies of the gr/gr deletion relative to the Italian and Spanish populations. (b) Schematic representation of the X chromosome depicting the two pseudoautosomal regions (PAR1 and PAR2) and the region including the recurrent deletions discovered by Giacco *et al.* [9]. On the right side, a model of the three deletions is provided with thinner lines indicating the deleted portions. In the upper part, the genomic size (Kb) of each deletion is specified. In the lower part, frequencies of each deletion in patients and controls are reported.

on its role as a significant genetic risk factor for impaired spermatogenesis [10[□]]. However, its routine screening is advised only in those countries where solid data exist on the level of risk conferred by this deletion. According to the guidelines [10[□]], testing for isolated gene-specific deletions in the AZFa and AZFb region is not advised because of their very low incidence as well as the highly heterogeneous testis phenotype.

Genetic counselling

AZF deletions are transmitted obligatory to the male offspring who will suffer from impaired sperm production. Apart from infertility, another potential risk is 45,XO/46XY mosaicism with ambiguous genitalia. Recently, it has been proposed that Y-chromosome microdeletions are also associated with a risk of developing pseudoautosomal region (PAR)–related anomalies. One study [12] reported that 5.4% of microdeletions carriers and normal karyotypes had *SHOX* haploinsufficiency in the PAR1, a condition related to disproportionate short stature

and diverse skeletal anomalies. This finding led to hypothesis that microdeletion carriers undergoing ART would expose their future sons to the risk of having not only fertility problems but also *SHOX*-linked pathologies. However, these alarming data have been contradicted by a more recent study performed on an almost doubled study population [13], in which none of the microdeletion carriers displayed *SHOX* deletion. This discrepancy might depend on either methodological issues or the different ethnic background of the two examined study populations.

CONGENITAL HYPOGONADOTROPIC HYPOGONADISM

Congenital hypogonadotropic hypogonadism (CHH) is a rare disease (incidence of 1:8000 men) [14] characterized by a deficit of gonadotropins leading to delayed puberty and azoospermia in men. Although major advances have been made for the discovery of novel candidate genes, the cause

remains unknown in about 50% of cases. cHH may be an isolated condition (normosmic cHH) or may be associated with hyposmia/anosmia (Kallmann syndrome). This classification is actually being questioned in recent years, because these conditions have been observed in different relatives belonging to the same family setting, as if they actually represented different phenotypic manifestations of the same genetic defect. Therefore, cHH is considered a complex genetic disease with variable expressivity, penetrance and inheritance fashions and does not seem to follow the rules of Mendelian inheritance. As such, the pathogenesis of cHH is likely to include the influence of environmental factors as well as the involvement of genetic variants in two or multiple interacting genes (oligo/digenia) [15]. A relatively newly identified feature of this disease is the presence of reversible cases, which represent about 10% of all cHH and does not seem to correlate with specific gene defects [14,16,17]. Mutations of cHH candidate genes are potentially related to late-onset hypogonadism and accordingly a *GNRHR* mutation has been described in this pathological condition [18].

Genetic counselling

cHH is highly responsive to gonadotrophin therapy; therefore, more than 90% of patients conceive either spontaneously or by assisted reproductive techniques. Genetic testing for the most frequently mutated candidate genes is available in many diagnostic genetic laboratories and depending on the gene involved, that is, whether a clear-cut cause–effect relationship is found, PGD or prenatal diagnosis may be offered to the couple.

PHARMACOGENETICS IN MALE INFERTILITY

Spermatogenic alterations might also be dependent on the presence of polymorphisms mapping to genetic regions that regulate the expression of genes involved in the reproductive endocrine system [19]. In this regard, single-nucleotide polymorphisms (SNPs) of the *FSHB* gene, encoding for the β -subunit of follicle-stimulating hormone (FSH), and the *FSHR* gene, encoding for the FSH receptor have been the object of recent studies in relationship with reproductive parameters and with individual responsivity to FSH therapy [20,21,22[■],23]. It is now widely accepted that one of the major determinants of serum FSH level is the *FSHB*-221G>T (rs10835638) in the gene promoter [20]. The T allele confers one-half of the activity of the wild-type promoter carrying the G allele [19]; therefore, levels of serum FSH, as well as sperm count, are significantly lower in G/T

heterozygous men and, even lower, in T/T homozygous carriers [20,21,22[■]]. Concerning the *FSHR* 2039A>G (rs6166) polymorphism and its effect on male reproductive parameters, a recent large study reported a highly significant effect of the G allele on testicular volume, which remarkably decreased in the G-allele carriers [23]. In another study, no significant associations but only trends of higher serum FSH and lower testicular volume have been found when this SNP was considered separately [22[■]]. However, in the same study a combined model involving both SNPs demonstrated that the coexistence of the two minor alleles (T/T and G/G) empowered the worst phenotype, that is, the lowest testicular volumes [22[■]]. Instead, the homozygous *FSHR* 2039 A/A allele, associated to higher receptor sensitivity, seems to compensate for the lower FSH serum levels conferred by the *FSHB* -221 T/T genotype. The advantage of detecting these polymorphic variants resides in the possibility of targeting them as a pharmacogenetic tool in the treatment of idiopathic infertile patients. This is of special importance if we consider the high cost of the therapy and that only a proportion of men (30–55%) result responsive to the treatment in terms of improvement of sperm parameters. To date only one pilot study addressed the question about the role of the *FSHB* -221 in relationship with FSH treatment [21]. In this study, all T/T homozygotes ($n=9$) resulted responsive to the treatment and showed a significantly higher improvement in sperm count and quality compared with carriers of the G allele [21]. In the group of G/T heterozygous, only 60% were responsive. Based on the Tüttelmann *et al.* study [22[■]], it is evident that future pharmacogenetic studies should be based on the combined analysis of the *FSHB* -221 G>T and the *FSHR* 2039 A>G, in order to further improve the selection of potentially responsive or unresponsive subjects.

MUTATIONS AND POLYMORPHISM IN THE ANDROGEN RECEPTOR

Pathogenic mutations in the gene encoding the androgen receptor (AR), located on Xq12, are associated with two main diseases: androgen insensitivity syndrome (AIS; MIM: 300068) and Kennedy syndrome (MIM: 313200). AIS can also manifest with a mild phenotype, for which patients suffer from oligo/azoospermia. The low frequency of AR mutations in OAT patients, the lack of selection criteria for testing and the extremely high number of mutations for which functional analyses should be performed in order to establish its cause–effect relationship with oligozoospermia limits the introduction of AR gene mutation screening in infertile

men. The *AR* gene also contains two polymorphic sites in the N-terminal *trans*-activation domain of the receptor: a polyglutamine tract $-(CAG)_n-$ and a polyglycine tract $-(GGC)_m$, which were the subject of many publications related to male infertility [24].

The $(CAG)_n$ length normally ranges between 6 and 39 repeats in the general population, with a median value that varies according to the ethnicity (21–22 in Caucasians, 19–20 in African-American, 22–23 in Asian populations). The originally described inverse relationship between CAG repeat length and the receptor *trans*-activation led to the hypothesis that longer CAG repeat conferred a higher risk for infertility and cryptorchidism. This hypothesis, which has been demonstrated also by an in-vivo animal model [25], has been questioned recently by novel functional [26] and observational studies reporting that both a longer CAG tract and a shorter CAG tract might have a negative effect on the receptor function [27,28]; hence, the highest transcription seemingly occurs in the presence of an optimal number of CAG repeats, which is represented by the median CAG length encountered in the general population. These models are based on genomic signaling pathways, but a recent study pioneered a new perspective: for instance, Davis-Dao *et al.* reported an unexpected association between shorter CAG repeats ($CAG \leq 19$) and the risk of cryptorchidism (an androgen-dependent disease), suggesting that the effect of this polymorphism might also be indirectly mediated by nongenomic signaling pathways [27]. We can speculate that the optimum range may vary between the genomic and nongenomic actions and also in different tissues because the effect of polyQ repeat on transactivation is cell specific, presumably due to distinct profiles of coregulator proteins [29]. Indeed, the role of CAG repeats in male infertility is probably more complex than it has been previously considered and more functional, and clinical studies are needed before this polymorphism can be introduced into the diagnostic setting.

NOVEL RESEARCH ASPECTS

Recently, whole-genome studies based on array-comparative genomic hybridization revealed the presence of a deletion burden in the genome (especially evident in the sex chromosomes) of infertile men [30–32]. This finding suggests that these men might display a higher genome instability, which might have consequences not only on their fertility status, but also on their general health. This phenomena could explain the lower life expectancy and higher morbidity of infertile men [33,34]. The

unexpectedly high number of X chromosome-linked genes with specific testicular expression (especially multicopy genes that have been recently acquired on the human X chromosome during evolution) [35] are in agreement with the higher deletion load in this chromosome in infertile men [31]. In 2013, for the first time, X chromosome-linked recurrent deletions have been reported and one of them (CNV67) resulted specific to oligo/azoospermia representing a novel diagnostic target in male infertility [9] (Fig. 1b).

CONCLUSION

In about 40% of male infertility cases, the pathophysiology remains unknown and the diagnosis and/or treatment still represent a challenge, especially when the infertile couple contemplates ART, and there exists the risk of transmitting genetic disorders to the future offspring.

Given that the so-called ‘idiopathic’ infertility cases are likely to be related to unidentified genetic/epigenetic and environmental factors, it is of outstanding importance to identify their missing cause. On the ‘genetic’ side, major progresses are expected with the diffusion of the Next-Generation Sequencing approach that will surely accelerate the identification of new genetic factors and will allow obtaining a comprehensive picture about the role of the combined action of low-size-effect genetic risk factors.

Acknowledgements

We gratefully acknowledge the Spanish Ministry of Health (grant FIS-11/02254) for financial support.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Krausz C. Male infertility: pathogenesis and clinical diagnosis. *Best Pract Res Clin Endocrinol Metab* 2011; 25:271–285.
2. Jungwirth A, Giwercman A, Tournaye H, *et al.* European Association of Urology guidelines on male infertility: the 2012 update. *Eur Urol* 2012; 62:324–332.
3. Dabaja AA, Schlegel PN. Microdissection testicular sperm extraction: an update. *Asian J Androl* 2013; 15:35–39.
4. Rives N, Milazzo JP, Perdrix A, *et al.* The feasibility of fertility preservation in adolescents with Klinefelter syndrome. *Hum Reprod* 2013; 28:1468–1479.
5. van Saen D, Gies I, De Schepper J, *et al.* Can pubertal boys with Klinefelter syndrome benefit from spermatogonial stem cell banking? *Hum Reprod* 2012; 27:323–330.
6. Gies I, De Schepper J, Van Saen D, *et al.* Failure of a combined clinical- and hormonal-based strategy to detect early spermatogenesis and retrieve spermatogonial stem cells in 47,XXY boys by single testicular biopsy. *Hum Reprod* 2012; 27:998–1004.

7. Harton GL, Tempest HG. Chromosomal disorders and male infertility. *Asian J Androl* 2012; 14:32–39.
 8. Groth K, Skakkebaek A, Høst C, *et al*. Clinical review: Klinefelter syndrome – a clinical update. *J Clin Endocrinol Metab* 2013; 98:20–30.
 9. Lo Giacco D, Chianese C, Ars E, *et al*. Recurrent X chromosome-linked deletions: discovery of new genetic factors in male infertility. *J Med Genet* 2013. [Epub ahead of print]
- For the first time in the literature, three novel recurrent deletions were identified on the X chromosome in association with spermatogenic impairment. Of particular note is CNV67 that, with its exclusive occurrence in azo/oligozoospermic patients, resembles the AZF regions on the Y chromosome with potential clinical implications.
10. Krausz C, Hoefsloot L, Simoni M, Tüttelmann F. EAA/EMQN best practice guidelines for molecular diagnosis of Y-chromosomal microdeletions: state-of-the-art 2013. *Andrology* 2014; 2:5–19.
- A comprehensive overview on all the relevant clinical and diagnostic issues related to the Y chromosome microdeletion screening. It states the importance of the analysis of deletion extension for predictive purposes and highlights the role of novel types of deletions.
11. Lo Giacco D, Chianese C, Sanchez-Curbelo J, *et al*. Clinical relevance of Y-linked CNV screening in male infertility: new insights based on the 8-year experience of a diagnostic genetic laboratory. *Eur J Hum Genet* 2013. [Epub ahead of print]
 12. Jorgez CJ, Weedon JW, Sahin A, *et al*. Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. *J Clin Endocrinol Metab* 2011; 96:E674–E679.
 13. Chianese C, Lo Giacco D, Tüttelmann F, *et al*. Y-chromosome microdeletions are not associated with SHOX haploinsufficiency. *Hum Reprod* 2013; 28: 3155–3160.
 14. Bonomi M, Libri DV, Guizzardi F, *et al*. New understandings of the genetic basis of isolated idiopathic central hypogonadism. *Asian J Androl* 2012; 14:49–56.
 15. Sykietis GP, Plummer L, Hughes VA, *et al*. Oligogenic basis of isolated gonadotropin-releasing hormone deficiency. *Proc Natl Acad Sci U S A* 2010; 107:15140–15144.
 16. Raivio T, Falardeau J, Dwyer A, *et al*. Reversal of idiopathic hypogonadotropic hypogonadism. *N Engl J Med* 2007; 357:863–873.
 17. Laitinen E-M, Tömmiska J, Sane T, *et al*. Reversible congenital hypogonadotropic hypogonadism in patients with CHD7, FGFR1 or GNRHR mutations. *PLoS One* 2012; 7:e39450.
 18. Tömmiska J, Jørgensen N, Christiansen P, *et al*. A homozygous R262Q mutation in the gonadotropin-releasing hormone receptor presenting as reversal of hypogonadotropic hypogonadism and late-onset hypogonadism. *Clin Endocrinol (Oxf)* 2013; 78:316–317.
 19. Hoogendoorn B, Coleman SL, Guy CA, *et al*. Functional analysis of human promoter polymorphisms. *Hum Mol Genet* 2003; 12:2249–2254.
 20. Grigorova M, Punab M, Zilaitienė B, *et al*. Genetically determined dosage of follicle-stimulating hormone (FSH) affects male reproductive parameters. *J Clin Endocrinol Metab* 2011; 96:E1534–E1541.
 21. Ferlin A, Vinanzi C, Selice R, *et al*. Toward a pharmacogenetic approach to male infertility: polymorphism of follicle-stimulating hormone beta-subunit promoter. *Fertil Steril* 2011; 96:1344–1349; e2.
 22. Tüttelmann F, Laan M, Grigorova M, *et al*. Combined effects of the variants ■ FSHB -211G>T and FSHR 2039A>G on male reproductive parameters. *J Clin Endocrinol Metab* 2012; 97:3639–3647.
- The first study in the literature that evaluated the combined effect of two SNPs on FSHB and FSHR demonstrating their influence of male reproductive fitness. A compensatory action between the two types of polymorphisms is also reported.
23. Grigorova M, Punab M, Poolamets O, *et al*. Study in 1790 Baltic men: FSHR Asn680Ser polymorphism affects total testes volume. *Andrology* 2013; 1:293–300.
 24. Davis-Dao CA, Tuazon ED, Sokol RZ, Cortessis VK. Male infertility and variation in CAG repeat length in the androgen receptor gene: a meta-analysis. *J Clin Endocrinol Metab* 2007; 92:4319–4326.
 25. Simanainen U, Brogley M, Gao YR, *et al*. Length of the human androgen receptor glutamine tract determines androgen sensitivity in vivo. *Mol Cell Endocrinol* 2011; 342:81–86.
 26. Nenonen H, Björk C, Skjaerpe P, *et al*. CAG repeat number is not inversely associated with androgen receptor activity in vitro. *Mol Hum Reprod* 2010; 16:153–157.
 27. Davis-Dao C, Koh CJ, Hardy BE, *et al*. Shorter androgen receptor CAG repeat ■ lengths associated with cryptorchidism risk among Hispanic white boys. *J Clin Endocrinol Metab* 2012; 97:E393–E399.
- The authors aimed at testing whether CAG repeat length polymorphism in exon 1 of the AR was associated with cryptorchidism. Surprisingly, the risk of cryptorchidism was significantly associated with shorter CAG repeats, leading to the hypothesis that androgens may influence the risk of cryptorchidism and related conditions through both genomic and nongenomic pathways.
28. Nenonen HA, Giwercman A, Hallengren E, Giwercman YL. Nonlinear association between androgen receptor CAG repeat length and risk of male subfertility – a meta-analysis. *Int J Androl* 2011; 34:327–332.
 29. Krausz C. An encore for the repeats: new insights into an old genetic variant. *J Clin Endocrinol Metab* 2012; 97:764–767.
 30. Tüttelmann F, Simoni M, Kliesch S, *et al*. Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PLoS One* 2011; 6:e19426.
 31. Krausz C, Giachini C, Lo Giacco D, *et al*. High resolution X chromosome-specific array-CGH detects new CNVs in infertile males. *PLoS One* 2012; 7:e44887.
 32. Lopes AM, Aston KI, Thompson E, *et al*. Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS Genet* 2013; 9:e1003349.
 33. Salonia A, Matloob R, Gallina A, *et al*. Are infertile men less healthy than fertile men? Results of a prospective case-control survey. *Eur Urol* 2009; 56:1025–1031.
 34. Jensen TK, Jacobsen R, Christensen K, *et al*. Good semen quality and life expectancy: a cohort study of 43,277 men. *Am J Epidemiol* 2009; 170:559–565.
 35. Mueller JL, Skaletsky H, Brown LG, *et al*. Independent specialization of the ■ human and mouse X chromosomes for the male germ line. *Nat Genet* 2013; 45:1083–1087.
- By improving the accuracy of the human X chromosome reference sequence, the authors provide an outstanding overview on the evolutionary processes that produce the gene repertoire on the human and mouse X chromosomes, describing how X-ampliconic genes, predominantly expressed in spermatogenic cells, specialize for sperm production.

SHORT REVIEW

The Y chromosome-linked copy number variations and male fertility

C. Krausz¹, C. Chianese¹, C. Giachini¹, E. Guarducci¹, I. Laface¹, and G. Forti²

¹Sexual Medicine and Andrology Unit; ²Endocrine Unit, Department of Clinical Physiopathology, University of Florence, Florence, Italy

ABSTRACT. Since the first definition of the AZoospermia Factor (AZF) regions, the Y chromosome has become an important target for studies aimed to identify genetic factors involved in male infertility. This chromosome is enriched with genes expressed exclusively or prevalently in the testis and their absence or reduction of their dosage is associated with spermatogenic impairment. Due to its peculiar structure, full of repeated homologous sequences, the Y chromosome is predisposed to structural rearrangements, especially deletions/duplications. This review discusses what is currently known about clinically relevant Y chromosome structural variations in male fertility, mainly focusing on copy number variations (CNVs). These CNVs include classical AZF deletions, gr/gr deletion and *TSPY1* CNV. AZF deletions are in a clear-cut cause-effect relationship with spermatogenic failure and they also

have a prognostic value for testis biopsy. gr/gr deletion represents the unique example in andrology of a proven genetic risk factor, providing an eight-fold increased risk for oligozoospermia in the Italian population. Studies on *TSPY1* CNV have opened new perspectives on the role of this gene in spermatogenic efficiency. Although studies on the Y chromosome have importantly contributed to the identification of new genetic causes and thus to the improvement of the diagnostic work-up for severe male factor infertility, there is still about 50% of infertile men in whom the etiology remains unknown. While searching for new genetic factors on other chromosomes, our work on the Y chromosome still needs to be completed, with special focus on the biological function of the Y genes. (J. Endocrinol. Invest. 34: 376-382, 2011)

©2011, Editrice Kurtis

INTRODUCTION

It has been known for many decades that the Y chromosome harbors the master gene for testis determination (SRY) and the so called AZoospermia Factor (AZF) regions, which contain genes involved in spermatogenesis. Structural anomalies, such as deletions and duplications of the AZF regions have been reported in association with male infertility and the screening for Y chromosome microdeletions became part of the routine diagnostic work-up of men with severe spermatogenic impairment (1). Recently, a role in spermatogenesis for a multicopy gene family, the *TSPY1* array, has also been proposed and it has been demonstrated that *TSPY1* copy number influences spermatogenic efficiency. In this review, we will discuss what is currently known about clinically relevant Y chromosome structural variations in male fertility, with special attention to copy number variations (CNV).

CNVs ON THE HUMAN Y CHROMOSOME: MECHANISM OF FORMATION

The Y chromosome is singular for its haploid nature which precludes recombination with the X-chromosome for most of its length. This has led to the consequent accumulation

of a high proportion of segmental duplications which provide the structural basis for the generation of CNVs (2, 3). The presence of such duplicated sequences allows two mechanisms to occur: a) gene conversion; b) non-allelic homologous recombination (NAHR). The first is an unidirectional conversion-based system of gene copy "correction" which permits the preservation of a certain number of Y genes from the gradual accumulation of deleterious mutations ensuring their continuity in time; on the other hand, NAHR produces recurrent deletions/duplications affecting the dosage of different Y genes (2, 4) (Fig. 1).

CNVs ON THE HUMAN Y CHROMOSOME: WHO ARE THEY?

Y-chromosomal microdeletions: the AZF deletions

Microdeletions of the Y chromosome are the most frequent known genetic cause of spermatogenic failure in infertile men, second only to the Klinefelter syndrome (5). The first association between azoospermia (absence of spermatozoa in the ejaculate) and microscopically detectable deletions of the long arm of the Y chromosome (Yq) has been demonstrated by Tiepolo and Zuffardi, in 1976 (6). They proposed the existence of a spermatogenesis factor, the AZF, encoded by a gene on distal Yq. With the development of molecular genetic tools it became possible to circumscribe the AZF region, in which microdeletions arise, and to highlight a certain deletion pattern with 3 recurrently deleted sub-regions in proximal, middle and distal Yq11, designated AZFa, AZFb and AZFc, respectively (7, 8) (Fig. 2).

Type of clinically relevant AZF deletions

The AZFa region is 792 Kb long and contains 2 single copy genes *USP9Y* and *DDX3Y* (former *DBY*) which are

Key-words: CNV, gr/gr deletion, male infertility, spermatogenesis, *TSPY*, Y chromosome.

Correspondence: C. Krausz, MD, Sexual Medicine and Andrology Unit, Department of Clinical Physiopathology, University of Florence, Viale Pieraccini 6, Florence 50139, Italy.

E-mail: c.krausz@dfc.unifi.it

Accepted February 3, 2011.

First published online March 21, 2011.

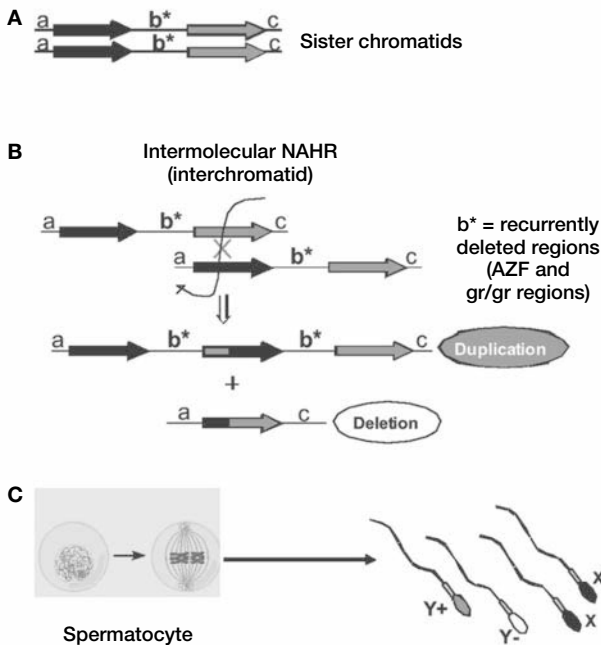


Fig. 1 - A) Representation of 2 sister chromatids on which the 4 arrows represent homologous sequences at the border of recurrently deleted regions on the Yq [i.e. AZoospermia Factor (AZF)a, AZFb, AZFc and gr/gr region). B) Schematic representation of the molecular mechanism responsible for deletions/duplications involving the AZF regions: interchromatidic non-allelic homologous Recombination (NAHR), occurring between homologous sequences orientated in the same direction. C) Meiotic division of a spermatocyte gives origin to four spermatozoa: i) two bearing the X-chromosome (X+); ii) two bearing the Y chromosome. If NAHR between sister chromatids occurs (see B), it will lead to a spermatozoon bearing: i) a Y chromosome with a duplication (Y+); ii) a Y chromosome with a deletion (Y-).

ubiquitously expressed. Complete AZFa deletions occur after homologous recombination between identical sequence blocks within the retroviral sequences in the same orientation HERVYq1 and HERVYq2 (9-11).

The complete deletion of AZFb is caused by homologous recombination between the palindromes P5/proximal P1 which removes also part of the AZFc region belonging to P1 (8). This deletion removes 6.2 Mb (including 32 copies of genes and transcription units). The AZFc region includes 12 genes and transcription units, each present in a variable number of copies making a total of 32 copies. The complete deletion of AZFc removes 3.5 Mb, originates from the homologous recombination between amplicons b2 and b4 in palindromes P3 and P1, respectively (12). Deletions of both AZFb and AZFc together occur in 2 breakpoints between P4/distal P1 (7.0 Mb, 38 gene copies removed) or between P5/distal P1 (7.7 Mb and 42 gene copies removed).

Clinical correlations of AZF deletions

The vast majority of complete AZF deletions are *de novo*, although exceptional cases of transmission have been reported and pertain uniquely the complete AZFc

deletion. However, "fertility" simply reflects that natural fertilization may occur even with relatively low sperm counts depending on the female partner's fertility status. AZF deletions are specific for spermatogenic failure as no deletions have been reported in the genomic DNA (derived from lymphocytes) of normozoospermic men (13). AZF deletions are likely to occur in germ cells during meiosis, when NAHR between sister chromatids may take place (Fig. 1). In order to investigate whether a testicular mosaicism for AZF deletions exists, we estimated the meiotic rate of AZFa deletions in 15 normozoospermic men. We found a meiotic deletion rate varying between 0.4 and 4.7×10^{-5} , implying that in an ejaculate containing several millions of spermatozoa, those bearing AZF deletions are several thousands (unpublished data).

Indications for AZF deletion screening are based on sperm count since clinical parameters such as hormone levels, testicular volume, varicocele, maldescended testis, infections, etc. do not have any predictive value (14-16). The test is currently performed in all infertile men with <5 millions/ml spermatozoa during the routine diagnostic work-up. The highest deletion frequency is found in idiopathic azoospermic men (10%) who are more likely to be carrier of genetic anomalies. AZF deletions are less frequent in severe oligozoospermic men (2-5%) and have been exceptionally reported in mild oligozoospermic men. Since AZF genes are mainly expressed in the testis, a number of studies have been undertaken in order to clarify if AZF deletions may cause testis related pathologies other than spermatogenic failure. No final evidence for a cause-effect relationship was observed with varicocele, cryptorchidism and testis cancer [for review see (5) and references therein].

Apart from the diagnostic value, Yq deletion screening provides additional prognostic information for testicular biopsy (TESE) in azoospermic men. In fact, deletions removing the entire AZFa or AZFb regions ("complete" deletions) are associated with Sertoli cell only syndrome (SCOS) and spermatogenic arrest, respectively, resulting in the absence of mature spermatozoa in the testis. Therefore, the presence of such deletions represents a negative predictive factor for TESE and carriers are discouraged from undergoing this invasive procedure (17, 18). Such a strict genotype/phenotype correlation is lacking for both partial deletions of these regions (exceptionally rare events) and AZFc deletions, which are associated with a semen phenotype varying from oligozoospermia to azoospermia with different testis histology. An azoospermic man with complete AZFc deletion has an average chance of 50% for successful testicular sperm retrieval. This variable phenotype might be due to a progressive regression of the germinal epithelium over time leading from oligozoospermia to azoospermia. An alternative explanation for this phenotypic variability could be the influence of the genetic background (i.e. compensation for the absence of Yq genes, by autosomal or X-linked factors), the presence of 45, XO lines (a more severe phenotype), and environmental factors in different individuals.

In case the deletion is found in a man undergoing ICSI or TESE/ICSI, genetic counselling is mandatory in order

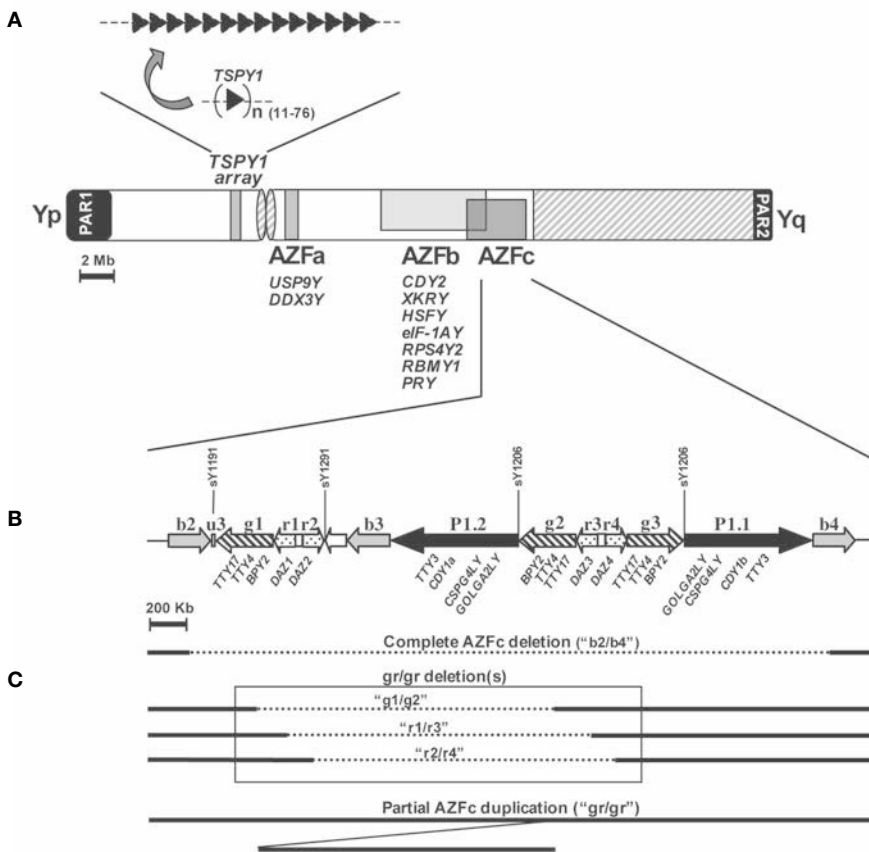


Fig. 2 - Schematic representation of the Y chromosome showing different regions/genes involved in spermatogenesis and Y-linked copy number variations. A) AZoospermia Factor (AZF)a, AZFb, and AZFc regions are located on the long arm of the Y chromosome (Yq) with an overlap between AZFb and AZFc; the TSPY1 gene is present on the short arm of the Y chromosome (Yp) arranged in a tandemly repeated array. B) AZFc region showing the location of multi-copy genes and transcription units in the reference sequence (Y haplogroup R). The arrows with the same motifs represent repeated homologous sequences, which may undergo NAHR. C) The "b2/b4" deletion (complete AZFc deletion) removing all AZFc genes is depicted. Three alternative breakpoints for gr/gr deletion(s) are shown which all remove half of the AZFc gene content. An example of partial AZFc duplication (gr/gr) is shown (similarly to the gr/gr deletion, different breakpoints may give origin to different types of gr/gr duplications).

to provide information about the risk of giving birth to a son with impaired spermatogenesis.

Diagnostic testing for AZF deletions

The diagnostic testing of Yq deletions should follow the European Academy of Andrology (EAA)/European Molecular Genetics Quality Network (EMQN) guidelines (1). The standard procedure is based on PCR amplification of AZF specific STS primers and control markers. Although alternative methods have been proposed (19), the use of the method described in the guidelines is highly advised for its high specificity and sensibility (detection of clinically relevant deletions is close to 100%). It is worth noting that the MSY sequence and the mechanism underlying microdeletions have definitely established that a fourth AZFd region postulated by Kent-First et al. (20) and considered in a popular commercial kit does not exist.

Partial AZFc deletions and duplications

The AZFc region consists almost entirely of repetitive sequence blocks called 'amplicons' which are arranged in direct and/or inverted repeats (12, 21). The region contains multicopy genes expressed specifically in the testis and their dosage may vary according to different types of rearrangements.

The first AZFc candidate gene isolated from the AZFc re-

gion on the long arm of the human Y chromosome was DAZ (Deleted in AZoospermia), which is specifically transcribed in the adult testis (22). The DAZ gene belongs to a gene family consisting of 3 members: *BOULE* on chromosome 2, *DAZ-Like (DAZL)* on chromosome 3 and *DAZ*, on the Yq. Members of this gene family are expressed exclusively in germ cells and encode testis-specific RNA-binding proteins that contain a highly conserved RNA-Recognition Motif (RRM) and a unique DAZ repeat (23). With regard to the reference sequence (corresponding to a Y chromosome belonging to haplogroup R), DAZ is present on the Y chromosome in 4 copies (DAZ1, DAZ2, DAZ3, and DAZ4). The AZFc region also harbors *CDY1*, present in 2 copies (*CDY1a* and *CDY1b*). *CDY* protein products have been identified as histone acetyltransferases with a strong preference for histone 4 (24), thus are likely to be involved in both spermatogenic histone replacement and DNA transcription. Other genes involved in AZFc deletions are *BPY2*, the function of which is still unknown, and 5 transcription units *TTY3*, *TTY4*, *TTY17*, *CSPG4LY*, and *GOLGA2LY*.

Due to its structure, the AZFc region is particularly susceptible to NAHR events which may cause the formation of both partial deletions or duplications and therefore alter the AZFc gene dosage (Fig. 2). Although a number of different partial AZFc deletions have been described, only one of them resulted to be clinically relevant. This is

the "gr/gr" deletion, named after the fluorescent probes ("green" and "red") used when it was detected for the first time (21). It removes half of the AZFc gene content, including two *DAZ* copies, one *CDY1* copy, and one *BPY2* copy. The clinical significance of the gr/gr deletion has been the object of a long debate. Controversies are mainly related to a number of selection biases (lack of ethnic/geographic matching of cases and controls; inappropriate selection of infertile and control men) and methodological issues (lack of confirmation of gene loss) (25-27). Moreover, another potential confounding factor derives from the fact that the frequency and phenotypic expression may vary among different ethnic groups, on the basis of the Y chromosome background; for example, in specific Y haplogroups, such as D2b, Q3, and Q1, common in Japan and certain areas of China, the deletion is fixed and apparently does not have any negative effect on spermatogenesis (28, 29). The presence of gr/gr deletion in Caucasian normozoospermic controls (although at a significantly lower frequency) prompted us to evaluate whether the Y background could influence the phenotypic variability in Caucasians, as well (30). It has been previously described that the loss of *DAZ1/DAZ2* and *CDY1* is prevalent (or even specific) in carriers with impaired sperm production (31-33) while it was hypothesized that the restoration of normal AZFc gene dosage in case of gr/gr deletion followed by b2/b4 duplication may explain the lack of effect on sperm count (22). Using a combined method based on gene dosage and gene copy definition of *DAZ* and *CDY1* genes (31), we could identify 4 different subtypes of gr/gr deletions characterized by the loss of different gene copies and could assess the presence of deletion followed by duplication. Notwithstanding the detailed characterization of subtypes of gr/gr deletions based on the type of missing gene copies and the detection of secondary rearrangements (deletion followed by b2/b4 duplication) together with the definition of Y haplogroups, it was impossible to define a specific pattern which would be associated with either a "neutral" or a "pathogenic" effect (30). Moreover, we also demonstrated that the restoration of normal gene dosage after secondary duplication is not specific for normozoospermic men. However, it is undeniable that the gr/gr deletion has some sort of effect even within the normal range of sperm count. It was observed, indeed, that normozoospermic carriers have a significantly lower sperm count, compared to men with intact Y chromosome (25). In addition, Yang et al. (34) reported that, in the Asian population, the deletion frequency drastically decreases in subgroups with sperm counts >50 millions spermatozoa/ml. More than 20 studies have been published during the last 7 years on this topic. According to the largest study population published to date on Caucasians (30), gr/gr deletion is significantly more frequent among oligozoospermic men (3.4%) compared to normozoospermic men (0.4%) and gr/gr deletion carriers are at a 7.9-fold increased risk for spermatogenic impairment [odds ratio (OR)=7.9 (95% confidence interval 1.8-33.8)]. As stated above, the heterogeneity of the study populations available in the literature thwarts the fulfillment of a reliable meta-analysis. Nevertheless, despite multiple biases, 4 meta-analyses

have been attempted on this topic all achieving significant OR reporting on average a 2-2.5-fold increase of risk (25, 35, 36). The gr/gr deletion represents a unique example of a significant risk factor for impaired sperm production.

The screening for gr/gr deletion is based on a PCR method described by Repping et al. (21). However, given a 5% false deletion rate detected in our multicenter study (30), deletions should be confirmed by gene dosage analysis.

The reasons why infertile men should undergo gr/gr deletion testing are mainly two: a) the deletion contributes to the etiopathogenesis of impaired sperm production since it is able to influence significantly the spermatogenic potential of the carrier; b) the couple should be aware that the deletion (i.e. a genetic risk factor for impaired sperm production) will be obligatorily transmitted to their male offspring and may become a complete AZFc deletion (i.e. a clear-cut causative factor for spermatogenic impairment) in the next generations (37, 38).

Since a detailed characterization of Y chromosomes belonging to different lineages found limited variation in the copy number of Y-linked genes, it raised the possibility of selective constraints (39). In this regard, about 90% of men carries four *DAZ* copies implying that a normal spermatogenesis requires an optimal copy number and therefore both a reduction and an increase of AZFc gene dosage may have a negative effect. This observation prompted two research groups to study the clinical consequences of partial AZFc duplications (26, 40), but they reached different conclusions, reporting an association between increased AZFc gene dosage and male infertility in the Han Chinese study and a lack of effect in our Italian study population. Since this discordance may reflect genuine ethnic differences, such as those observed for the corresponding partial AZFc deletions, if increased AZFc gene content is to play a role in spermatogenic impairment, the effect will probably be modulated by population specific factors. Further studies are needed to provide evidence to support this hypothesis.

AZF gene-specific deletions

Despite the efforts of many laboratories, only 5 cases of confirmed (after sequencing the breakpoints) isolated Yq gene mutation have been reported to date (41). The rarity of single AZF gene-specific mutations or deletions is in sharp contrast with the relatively high frequency of AZF deletions (described above) and this might be a consequence of the peculiar organization of the Y chromosome, which makes it more prone to the loss of large portions – such as the AZF region – rather than single genes. The only reported isolated mutation occurs in the AZFa region which contains two widely expressed genes, *USP9Y* and *DDX3Y* (2). In the first place, sequencing of the two genes in 576 patients brought to believe that the loss of *USP9Y* had a direct effect on spermatogenesis, causing azoospermia, whereas no mutation was found in the *DDX3Y* gene (42). However, following findings revealed that what seemed to be a definitive result turned out to be just one of the possible phenotypes related to *USP9Y* deletion. In fact, Luddi et al. (43) reported that

this deletion has no effect on spermatogenesis and is thus compatible with fertility. On the other hand, previous studies (42, 44, 45) irrefutably demonstrate that the loss of the gene disturbs spermatogenesis to different degrees and that natural transmission is possible in case of a mild phenotype. Therefore, *USP9Y* has been proposed as a fine spermatogenic modulator (44), the absence of which is compatible with a highly variable phenotype probably linked to the genetic or other background of the carrier (41). Given the extreme rarity of AZF gene specific deletions and the heterogeneous phenotype of the *USP9Y* deletion, the routine screening for AZF gene specific deletions is not advised.

TSPY1 array

During the last years, Y-linked CNV analyses have been extended to the short arm of the Y chromosome which contains a *TSPY1* gene array with variable number of *TSPY1* copies (46, 47 and references therein). The *TSPY1* belongs to a protein superfamily comprising SET and NAP, which are activating factors of the replication process. Indeed, *TSPY1* is abundantly expressed in early stages of tumorigenesis in gonadoblastoma and could be potentially involved in other human cancers (48). Expression analysis in the testis indicates the involvement of the *TSPY1* in spermatogenesis as a pro-proliferative factor (48). In fact, *TSPY1* is mainly expressed in gonocytes/pre-spermatogonia of embryonic testis and in spermatogonia and spermatocytes at meiotic prophase I in adult testis. A role in early fetal germ cells development has also been addressed by Schoner et al. (49) who provided evidence of *TSPY1* ability to partially rescue spermatogenesis and fertility in transgenic Kit^{W-v}/Kit^{W-v} mice. *TSPY1* is unusual in being arranged in a tandem array of 20.4 Kb of repeated units, bearing a single active *TSPY1* copy each (Fig. 2). Although copy number varies among individuals within a range of 11 to 76 (26, 46, 50, 51), the majority of men (about 65% of the Italian population) remain within a restricted interval (21 to 35 copies) (46). The evolutionary conservation of multiple *TSPY1* copies on the Y chromosome of other mammals as well as the above mentioned limited variation in copy number in humans suggest that a minimum *TSPY1* copy number is likely to be maintained through selection (52). Only few studies have focused on the eventual *TSPY1* influence on spermatogenesis and frustratingly they all reached 3 different conclusions, probably due to study design biases (46, 50, 51). Indeed, crucial for a reliable analysis is the *TSPY1* CNV susceptibility to stratification biases. As a matter of fact, significantly different means of *TSPY1* copy number were found among different Y haplogroups (46, 53), highlighting the importance of Y haplogroups-matching between cases and controls. The only available study to date in which cases and controls were matched for Y hgr distribution has been performed in the Italian population by our group. The method used for the detection of *TSPY1* copy number was validated against pulsed-field gel electrophoresis (the gold standard method) (46). The initially published study population has been recently enlarged and previous results confirmed i.e. a significantly lower *TSPY1* copy number in 212 infertile men with abnormal sperm parameters compared to 168 normo-

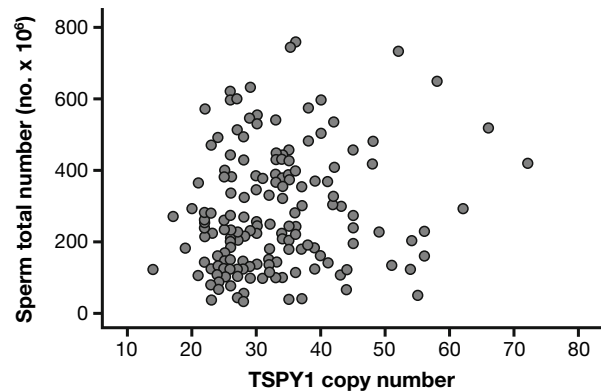


Fig. 3 - Scatter plots between *TSPY1* copy number and total sperm count in normozoospermic controls. Spearman's correlation coefficient (Rho) = 0.179; p = 0.021.

zoospermic subjects (28.5 ± 7.9 vs 32.6 ± 10.1 , respectively; $p < 0.001$) has been found. The relevance of *TSPY1* CNV in spermatogenesis is also attested by the positive correlation observed with sperm count both in infertile and normozoospermic subjects (Fig. 3). In the light of these findings, low *TSPY1* copy number can be regarded as a new genetic risk factor for male infertility with potential clinical consequences and should be taken into consideration in the context of a multigenic approach to idiopathic infertility.

CONCLUSIVE REMARKS

The pivotal role of the Y chromosome in spermatogenesis is supported by the presence of Y-linked genes specifically expressed in germ cells and by pathological phenotypes deriving from the deletion of regions containing the genes mentioned above. Although Y studies have importantly contributed to the identification of new genetic factors in male infertility and thus to the improvement of the diagnostic work-up of severe male factor infertility, there are still many unsolved issues. Among them the most relevant are: a) the lack of knowledge about the exact function of AZF gene products; b) the correlation, if existing, between *TSPY1* copy number and its level of expression; c) lack of information about the consequences of AZFc gene dosage variation on mRNA involved in spermatogenesis and eventually in embryogenesis. It is also unknown whether the "fragility" of the Y chromosome is a marker for general "genomic instability" potentially affecting the general health status of the Y deletion carrier. In this regard, gr/gr deletion has been reported as a risk factor also for testicular germ cell tumors, but data need further confirmation (54). Beside the Y chromosome-linked genes, several thousands autosomal and X-linked genes are predicted to play a role in the complex process of spermatogenesis. The two sex chromosomes share common features, in particular the peculiar repeated structure containing a number of multicopy gene families with testis specific expression. Therefore, we expect that similarly to the Y chromosome, also X-linked CNV

would affect gene dosage and thus be responsible for a portion of severe male factor infertility. Then, it seems high time to stop focusing only on the Y chromosome rearrangements and to start shifting our attention also on its partner, the X chromosome.

ACKNOWLEDGMENTS

Funding from telethon Italy (no. GGP08204; to C.K.) and PRIN 2007-2009 (to C.K.) are acknowledged.

REFERENCES

- Simoni M, Bakker E, Krausz C. EAA/EMQN best practice guidelines for molecular diagnosis of y-chromosomal microdeletions. State of the art 2004. *Int J Androl* 2004, 27: 240-9.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003, 423: 825-37.
- Jobling MA. Copy number variation on the human Y chromosome. *Cytogenet Genome Res* 2008, 123: 253-62.
- Turner DJ, Miretti M, Rajan D, et al. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet* 2008, 40: 90-5.
- Krausz C, Degl'Innocenti S. Y chromosome and male infertility: update, 2006. *Front Biosci* 2006, 11: 3049-61.
- Tiepolo L, Zuffardi O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet* 1976, 34: 119-24.
- Vogt PH, Edelmann A, Kirsch S, et al. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Mol Genet* 1996, 5: 933-43.
- Repping S, Skaletsky H, Lange J, et al. Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am J Hum Genet* 2002, 71: 906-22.
- Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME. Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J Med Genet* 2000, 37: 752-8.
- Kamp C, Hirschmann P, Voss H, Huellen K, Vogt PH. Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Hum Mol Genet* 2000, 9: 2563-72.
- Sun C, Skaletsky H, Rozen S, et al. Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum Mol Genet* 2000, 9: 2291-6.
- Kuroda-Kawaguchi T, Skaletsky H, Brown LG, et al. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* 2001, 29: 279-86.
- Krausz C, Forti G, McElreavey K. The Y chromosome and male fertility and infertility. *Int J Androl* 2003, 26: 70-5.
- Maurer B, Gromoll J, Simoni M, Nieschlag E. Prevalence of Y chromosome microdeletions in infertile men who consulted a tertiary care medical centre: the Munster experience. *Andrologia* 2001, 33: 27-33.
- Oates RD, Silber S, Brown LG, Page DC. Clinical characterization of 42 oligospermic or azospermic men with microdeletion of the AZFc region of the Y chromosome, and of 18 children conceived via ICSI. *Hum Reprod* 2002, 17: 2813-24.
- Frydelund-Larsen L, Krausz C, Leffers H, et al. Inhibin B: a marker for the functional state of the seminiferous epithelium in patients with azoospermia factor C microdeletions. *J Clin Endocrinol Metab* 2002, 87: 5618-24.
- Brandell RA, Mielnik A, Liotta D, et al. AZFb deletions predict the absence of spermatozoa with testicular sperm extraction: preliminary report of a prognostic genetic test. *Hum Reprod* 1998, 13: 2812-5.
- Krausz C, Quintana-Murci L, McElreavey K. Prognostic value of Y deletion analysis: what is the clinical prognostic value of Y chromosome microdeletion analysis? *Hum Reprod* 2000, 15: 1431-4.
- Fattoruso O, Zarrilli S, Coto I, De Rosa M, Lombardi G, Sacchetti L. Prevalence of Y microdeletions in azospermic and severe oligozoospermic men in Southern Italy: application of a rapid capillary electrophoresis method. *J Endocrinol Invest* 2009, 32: 223-7.
- Kent-First M, Muallem A, Shultz J, et al. Defining regions of the Y-chromosome responsible for male infertility and identification of a fourth AZF region (AZFd) by Y-chromosome microdeletion detection. *Mol Reprod Dev* 1999, 53: 27-41.
- Repping S, Skaletsky H, Brown L, et al. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet* 2003, 35: 247-51.
- Reijo R, Lee TY, Salo P, et al. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet* 1995, 10: 383-93.
- Yen PH. Putative biological functions of the DAZ family. *Int J Androl* 2004, 27: 125-9.
- Lahn BT, Tang ZL, Zhou J, et al. Previously uncharacterized histone acetyltransferases implicated in mammalian spermatogenesis. *Proc Natl Acad Sci USA* 2002, 99: 8707-12.
- Visser L, Westerveld GH, Korver CM, et al. Y chromosome gr/gr deletions are a risk factor for low semen quality. *Hum Reprod* 2009, 24: 2667-73.
- Giachini C, Laface I, Guarducci E, Balercia G, Forti G, Krausz C. Partial AZFc deletions and duplications: clinical correlates in the Italian population. *Hum Genet* 2008, 124: 399-410.
- Navarro-Costa P, Goncalves J, Plancha CE. The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility. *Hum Reprod Update* 2010, 16: 525-42.
- Yang Y, Ma M, Li L, et al. Differential effect of specific gr/gr deletion subtypes on spermatogenesis in the Chinese Han population. *Int J Androl* 33: 745-54.
- Sin HS, Koh E, Shigehara K, et al. Features of constitutive gr/gr deletion in a Japanese population. *Hum Reprod* 2010, 25: 2396-403.
- Krausz C, Giachini C, Xue Y, et al. Phenotypic variation within European carriers of the Y-chromosomal gr/gr deletion is independent of Y-chromosomal background. *J Med Genet* 2009, 46: 21-31.
- Giachini C, Guarducci E, Longepied G, et al. The gr/gr deletion(s): a new genetic test in male infertility? *J Med Genet* 2005, 42: 497-502.
- Fernandes S, Huellen K, Goncalves J, et al. High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia. *Mol Hum Reprod* 2002, 8: 286-98.
- Ferlin A, Tessari A, Ganz F, et al. Association of partial AZFc region deletions with spermatogenic impairment and male infertility. *J Med Genet* 2005, 42: 209-13.
- Yang Y, Xiao CY, A ZC, Zhang SZ, Li X, Zhang SX. DAZ1/DAZ2 cluster deletion mediated by gr/gr recombination per se may not be sufficient for spermatogenesis impairment: a study of Chinese normozoospermic men. *Asian J Androl* 2006, 8: 183-7.
- Tuttelmann F, Rajpert-De Meyts E, Nieschlag E, Simoni M. Gene polymorphisms and male infertility--a meta-analysis and literature review. *Reprod Biomed Online* 2007, 15: 643-58.
- Stouffs K, Lissens W, Tournaye H, Haentjens P. What about gr/gr deletions and male infertility? Systematic review and meta-analysis. *Hum Reprod Update* 2011, 17: 197-209.
- Zhang F, Lu C, Li Z, et al. Partial deletions are associated with an increased risk of complete deletion in AZFc: a new insight into the role of partial AZFc deletions in male infertility. *J Med Genet* 2007, 44: 437-44.
- Lu C, Zhang J, Li Y, et al. The b2/b3 subdeletion shows higher risk of spermatogenic failure and higher frequency of complete AZFc deletion than the gr/gr subdeletion in a Chinese population. *Hum Mol Genet* 2009, 18: 1122-30.
- Repping S, van Daalen SK, Brown LG, et al. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet* 2006, 38: 463-7.
- Lin YW, Hsu LC, Kuo PL, et al. Partial duplication at AZFc on the Y chromosome is a risk factor for impaired spermatogenesis in Han Chinese in Taiwan. *Hum Mutat* 2007, 28: 486-94.
- Tyler-Smith C, Krausz C. The will-o'-the-wisp of genetics--hunting for the azoospermia factor gene. *N Engl J Med* 2009, 360: 925-7.

42. Sun C, Skaletsky H, Birren B, et al. An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. *Nat Genet* 1999,23: 429-32.
43. Luddi A, Margollicci M, Gambera L, et al. Spermatogenesis in a man with complete deletion of USP9Y. *N Engl J Med* 2009, 360: 881-5.
44. Krausz C, Degl'Innocenti S, Nuti F, et al. Natural transmission of USP9Y gene mutations: a new perspective on the role of AZFa genes in male fertility. *Hum Mol Genet* 2006, 15: 2673-81.
45. Brown GM, Furlong RA, Sargent CA, et al. Characterisation of the coding sequence and fine mapping of the human DFFRY gene and comparative expression analysis and mapping to the Sxrb interval of the mouse Y chromosome of the Dffry gene. *Hum Mol Genet* 1998, 7: 97-107.
46. Giachini C, Nuti F, Turner DJ, et al. TSPY1 copy number variation influences spermatogenesis and shows differences among Y lineages. *J Clinical Endocrinol Metab* 2009, 94: 4016-22.
47. Krausz C, Giachini C, Forti G. TSPY and Male Fertility. *Genes* 2010, 1: 308-16.
48. Lau YF, Li Y, Kido T. Gonadoblastoma locus and the TSPY gene on the human Y chromosome. *Birth Defects Res C Embryo Today* 2009, 87: 114-22.
49. Schöner A, Adham I, Mauceri G, et al. Partial rescue of the KIT-deficient testicular phenotype in KitW-v/KitW-v Tg (TSPY) mice. *Biol Reprod* 83: 20-6.
50. Vodicka R, Vrtel R, Dusek L, et al. TSPY gene copy number as a potential new risk factor for male infertility. *Reprod Biomed Online* 2007, 14: 579-87.
51. Nickkholgh B, Noordam MJ, Hovingh SE, van Pelt AM, van der Veen F, Repping S. Y chromosome TSPY copy numbers and semen quality. *Fertil Steril* 2010, 94: 1744-7.
52. Tyler-Smith C. An evolutionary perspective on Y-chromosomal variation and male infertility. *Int J Androl* 2008, 31: 376-82.
53. Mathias N, Bayés M, Tyler-Smith C. Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 1994, 3: 115-23.
54. Nathanson KL, Kanetsky PA, Hawes R, et al. The Y deletion gr/gr and susceptibility to testicular germ cell tumor. *Am J Hum Genet* 2005, 77: 1034-43.

Chapter 2

Genomic Changes in Spermatozoa of the Aging Male

Chiara Chianese, Sara Brilli, and Csilla Krausz

Abstract Modern society is witnessing a widespread tendency to postpone parenthood due to a number of socioeconomic factors. This ever-increasing trend relates to both women and men and raises many concerns about the risks and consequences lying beneath the natural process of aging. The negative influence of the advanced maternal age has been thoroughly demonstrated, while the paternal age has attracted comparatively less attention. A problematic issue of defining whether advanced paternal age can be considered an independent risk factor, not only for a man's fertility but also for the offspring's health, is represented by the difficulty, linked to reproductive studies, in characterizing the impact of maternal and paternal age, separately. Researchers are now trying to overcome this obstacle by directly analyzing the male germ cell, and emerging data prove this sperm-specific approach to be a valid tool for providing novel insights on the effects of aging on the spermatozoa and, thus, on the reproductive outcome of an aging male. The purpose of this chapter is to summarize most of what is known about the relationship between male aging and changes in the spermatozoa, giving special focus on the events occurring with age at the genomic level.

Keywords Aging male • Spermatozoa • Genomic anomalies

Introduction

History provides fascinating episodes of men fathering children at very old ages. In 1935, a 94-year-old man was the oldest age-of-paternity case reported by a scientific publication (Seymour et al. 1935). Other examples of greatly aged fathers

C. Chianese • S. Brilli • C. Krausz (✉)

Andrology Unit, Department of Experimental and Clinical Biomedical Sciences,
University of Florence, Florence, Italy
e-mail: csillagabriella.krausz@unifi.it

have appeared more recently, such as that of two Indian farmers, Nanu Ram Jogi, who fathered his twenty-first child at age 91 in 2007, and Ramjit Raghav, who became the world's oldest dad, giving birth to his first child at the age of 94.

Beyond anecdotal curiosity, it is interesting to note that in industrialized countries delayed parenthood is becoming an increasingly widespread social phenomenon. Many factors account for this trend, but referring to literature citations increasing life expectancy, economic stability, and career ambitions represent the most relevant issues for parenthood postponement, raising not a few concerns about age-associated risks and consequences. Indeed, the process of aging can be ascribed to a number of endogenous and environmental factors inducing DNA damage.

As demonstrated by epidemiological data, the last decades saw a considerable increase in the mean age of childbearing mothers, reflecting for women a birth rate shift toward 35 years of age and older. A similar tendency shows up in the paternity rate, which has been continuously rising since 1980, in parallel with a decrement of paternity among men between 25 and 29 years old (Martin et al. 2007). In fact, the percentage of men fathering children over 35 years old has been markedly rising from 15 % to 25 % during the last 40 years, as has the number of men aged between 50 and 54 desiring to conceive children (Fisch 2009). In line with these data, a rise in the number of fathers over 60 is predicted to happen over the next 10–20 years (US Census Bureau 2005).

The effect of delayed motherhood has been studied to such an extent that it is now possible to acknowledge advanced maternal age as the most important risk factor for infertility, spontaneous abortions, and genetic defects among offspring. Only lately has male age attracted more attention in this regard, but whether a comparable age-dependent effect also exists for delayed fatherhood remains to be fully elucidated. There is suggestive epidemiological evidence that paternal age correlates with an increased incidence of abnormal reproductive outcomes and heritable defects (Tarin et al. 1998; de la Rochebrochard and Thonneau 2002), including several types of genomic modification. In particular, there is growing evidence that advancing male age is associated with an increased frequency of certain genetic and chromosomal defects in spermatozoa (Crow 2000; Shi and Martin 2000; Tiemann-Boege et al. 2002; Bosch et al. 2003; Slotter et al. 2004). The last-named authors prove that the male germ cell represents a direct target of study for a straightforward identification of merely paternal risk factors, overcoming the difficulty, often present in reproductive studies, in distinguishing between the impact of maternal and paternal ages. Nevertheless, it is as yet impossible to even define a general consensus of what can be considered *advanced* paternal age, although some studies and the precautionary measures taken for sperm donations destined for assisted conception advise a threshold of 40 years old (de la Rochebrochard et al. 2003).

This chapter aims at summarizing data available on male age-related effects in spermatozoa, describing with specific focus how the sperm genome can be modified during meiosis of aging men.

Chromosomal Alterations in the Aging Male

Chromosomal aberrations were among the first observed manifestations of decreased genome integrity with age. Studies comprising chromosomal analyses from human peripheral blood lymphocytes corroborate these observations, demonstrating that the occurrence of various genomic changes (i.e., aneuploidies, translocations, acentric fragments, chromosomal loss, micronuclei formation) increases linearly with the age of the individuals (Sloter et al. 2004). In light of these data, concerns have arisen that certain types of chromosomal alterations may increase with age in male germ cells as well.

Aneuploidies

In humans, aneuploidies represent the most common heritable chromosomal anomaly, with approximately 0.3 % of infants born with an altered number of chromosomes (Hassold 2001). This genomic rearrangement derives mainly from nondisjunction during meiotic divisions [both meiosis I (MI) and meiosis II (MII)] and its primary reproductive consequence is spontaneous abortion. Analyses on fetal material retrieved from abortuses show that 60 % of all aneuploidies consist in 45, XO monosomy and trisomies of chromosomes 16, 21, and 22 (Hassold 2001). A subset of aneuploidies involving sex chromosomes as well as trisomies 13, 21, and 18 make it to birth and lead to typical developmental and morphological defects.

Several studies proved maternal age to be a risk factor for trisomy formation, although the nature of such a relationship is not unequivocally definable: in fact, the effect of maternal age is differently exerted among chromosomes, increasing either linearly (e.g., trisomy 16) or exponentially (e.g., trisomy 21). The characterization of such an age-dependent connection for individual trisomies has not progressed much for men due to the relatively low number of trisomy cases determined to be paternally derived. Moreover, whether paternal age is associated with the generation of aneuploidies is still controversial due to the paucity of data available on the affected offspring because most abnormal embryos are lost and because studies performed so far were limited by the difficulty in separating paternal from maternal effects (Wiener-Megnazi et al. 2012). Cytogenetic data from human oocytes, fertilized eggs, preimplantation embryos, and spermatozoa allowed to compensate for the issue of discriminating the pure paternal effect. The approach of targeting solely the germ cell genome, extracting parent-specific information, indicates that most constitutional aneuploidies are generated *de novo* during parental meiosis.

The first data about the chromosomal content of human spermatozoa date back to 1978 and derive from the insemination of hamster eggs with human spermatozoa, a cytogenetic method known as the hamster-egg penetration test (Rudak et al. 1978).

This study showed that 2–3% of spermatozoa from normal men carried an aneuploid karyotype. Other large hamster-egg studies followed, but an effect of the donor's age on the rate of aneuploidies was not found (Estop et al. 1995; Brandriff et al. 1988); this was probably due to a bias related to the study population, which included only a few men aged over 40 years old. Finally, two more studies based on the same method reached divergent conclusions: in one, a significantly negative correlation was found between the rate of aneuploidy and the donor's age (Martin and Rademaker 1987), while in the other study the authors observed a significantly higher incidence of hyperploid spermatozoa from a comparison between seven old donors (aged between 59 and 74 years old) and five young donors (aged between 23 and 39 years old) (Sartorelli et al. 2001).

In the early 1990s, fluorescent in situ hybridization (FISH) replaced the hamster-egg method for the detection of sperm aneuploidy, introducing the advantage of a quicker labour- and cost-saving analysis of greater amounts of spermatozoa. With regard to sperm autosomal aneuploidies, FISH data show modest evidence for a paternal age effect. One study (Martin et al. 1995) reported an increase with age in disomy 1 in the spermatozoa of men aged from 21 to 52 years old, although none of the following studies confirmed their result. Another study found a significant correlation between a decreased incidence of chromosome 18 sperm disomy and increased age (Robbins et al. 1997). In contrast, studies on chromosome 21 sperm disomy all converge on the independence of such a rearrangement from men's age, with the exception of only one small study (Rousseaux et al. 1998) in which the authors found that advanced age correlated with a higher incidence of sperm disomy 21.

A different scenario is offered by FISH studies on sex chromosomes, for which more distinct evidence exists for an age-related increase in aneuploidies in male germ cells. In the literature, 11 sperm FISH studies are available on the effects of paternal age on the frequency of aneuploidy formation within spermatid X and Y chromosomes, and only 2 reached a negative conclusion with respect to the age–aneuploidy link. The rest provide evidence that errors in MI and MII are more likely to occur with the advancement of age. As for nondisjunctions in both MI and MII, the literature reports a general positive paternal age effect by which men aged over 50 have about a two- to threefold higher risk of carrying spermatozoa with a 24, XY karyotype as a consequence of MI errors (Guttenbach et al. 2000; Bosch et al. 2001; Asada et al. 2000; Griffin et al. 1995; Lowe et al. 2001) and a two to threefold higher frequency of producing X or Y disomic spermatozoa as a consequence of MII errors (Kinakin et al. 1997; Martin et al. 1995; Robbins et al. 1997; Rubes et al. 1998).

Multiprobe FISH analyses have proven their utility in the detection of sperm diploidy, the incidence of which in association with paternal aging is still controversial. For instance, the literature offers a number of studies reporting an age-related increase of the frequency of diploid sperm in older men compared to younger men with about a twofold increased risk. However, several other studies do not reach the same conclusion because they report no association between sperm disomy formation and advanced paternal age (Sloter et al. 2004).

Structural Aberrations

Although the incidence of structural chromosomal abnormalities is lower compared to aneuploidies at birth (0.25 % versus 0.33 %, respectively) (Hassold 1998), a study based on chromosome heteromorphisms first estimated that 80 % of such de novo rearrangements are of paternal origin (Olson and Magenis 1988). These data were supported by subsequent findings on paternally derived aberrations. Thomas et al. observed a paternal derivation of de novo unbalanced structural chromosomal abnormalities detectable by light microscopy, with 84 % interstitial deletions and 58 % duplications and rings (Thomas et al. 2006). Even more recently, de novo microdeletions associated with de novo reciprocal translocations as well as cases of complex chromosomal rearrangements were determined to be paternally derived in all cases. Likewise, array comparative genomic hybridization (aCGH) analyses helped in determining that all de novo deletions identified in men carrying balanced translocations and abnormal phenotypes derived from the fathers (Baptista et al. 2008). Other studies reported that both a recurrent de novo translocation, i.e., t(11;22), and nonrecurrent balanced reciprocal translocations were almost entirely of paternal origin, with 100 % for the former and 96 % for the latter being inherited from the fathers (Kurahashi et al. 2009; Ohye et al. 2010; Thomas et al. 2012).

Concerning the aging male, the literature provides conflicting evidence of a paternal age effect for structural rearrangements. The incongruence emerges between case studies noting that structural aberrations spontaneously occur in children conceived by older fathers and population-based studies in which such a correlation does not appear to be real (Sloter et al. 2004). Although the paternal contribution still seems rather high, information on structural aberrations in human male gametes is still scarce. This is partly due to the overall lower occurrence of such rearrangements among live births that render the paternal effect enormously complicated to define. However, the mounting development of assays that allow the detection of structural chromosomal aberrations directly within spermatozoa represents an important incentive for the evaluation of those factors, such as age, that will potentially increase the formation rate of anomalies in a man's sperm population.

The first clue of a relationship between the incidence of structural aberrations in male germ cells and paternal age comes from rodent studies. There is consistent evidence that structural aberrations in rodent spermatozoa increase with age, although the pre- and postmeiotic spermatogenetic compartments seem differently affected, with late-step spermatids, and not primary spermatocytes, showing a greater fold increase in the frequency of abnormalities between old and young mice (Pacchierotti et al. 1983). These data are supported by two micronucleus assays (Allen et al. 1996; Lowe et al. 1995) focusing on the age effect on the frequency of aberrations in round spermatids of mice and hamsters, respectively, and both leading to the conclusion that older animals have a significantly higher frequency of unstable aberrations in their spermatids compared to young animals.

Concerning human semen, the hamster-egg method in the first instance revealed itself as a relevant tool for the detection of spermatozoa bearing structural

chromosomal abnormalities such as unrejoined breaks and acentric fragments, of which 75 % resulted in unstable aberrations. The examination of 1,582 sperm chromosomal complements from 30 fertile men divided into six age groups ranging from 20 to 24 years to older than 45 years reported a fourfold increase in the total structural chromosomal abnormalities for older men (Martin and Rademaker 1987). The reanalysis of these data by Slotter et al. (2004) demonstrates that this effect is mainly due to the significant increase in chromosomal breaks, but not in acentric fragments, indicating the greater susceptibility to aging of postmeiotic DNA-repair-free spermatids. Another human-sperm/hamster-egg study (Sartorelli et al. 2001), including several men between 59 and 74 years old, reported a significantly higher frequency of acentric fragments and of complex radial figures in sperm complements of older donors compared to younger donors.

Notwithstanding its importance in producing the aforementioned results, the hamster-egg method is inefficient to measure the frequency of deletions and duplications as well as of the so-called stable rearrangements, i.e., translocations, inversions, insertions, isochromosomes, small deletions, and small duplications, in the spermatozoa and thus has been replaced by the FISH strategy. An age-related effect was observed for the frequency of centromeric deletions of chromosome 1 in a cohort of 18 men aged 20–58 years old (McInnes et al. 1998); likewise, a significant age-related increase was reported for the frequency of spermatozoa with duplications and deletions at the centromeric and subtelomeric regions of chromosome 9 in a cohort of 18 men aged 24–74 years old (Bosch et al. 2003). Another FISH-based analysis demonstrated a significant increase in the frequency of spermatozoa carrying breaks and segmental duplications and deletions of chromosome 1 among older men compared to younger men. In particular, older men showed twice the frequency of segmental duplications and deletions in chromosome 1 in their spermatozoa. Similarly, the researchers found a significant age-related increase in the frequency of spermatozoa carrying breaks within the 1q12 fragile-site region that was almost doubled in older men (Slotter et al. 2007). A more recent study based on a multi-color, multichromosome FISH strategy was performed on the semen of ten male donors 23–74 years old and found that older patients had a higher rate of structural abnormalities (6.6 %) compared to younger men (4.9 %); interestingly, although both duplications and deletions occurred more frequently in older men, an excess of duplications versus deletions was observed in both groups. In addition, the authors demonstrated a nonlinear distribution of duplications and deletions along the chromosomes and observed an inclination toward a higher susceptibility to rearrangements in larger chromosomes (Templado et al. 2011).

Overall, both human and animal studies suggest that the increased trend of delaying fatherhood could predict an augmented risk of delivering offspring liable to paternally derived genetic diseases resulting from chromosomal aneuploidies or structural aberrations, assuming that spermatozoa bearing such rearrangements are as capable of fertilizing as normal spermatozoa. However, animal models provide evidence for a paternal age effect mostly on chromosomal breaks, duplications, and deletions rather than chromosomal numeric alterations. Along these lines, duplications appear to occur more frequently than deletions, suggesting a mitotic rather than meiotic

origin for some of these sperm de novo abnormalities. As for stable rearrangements, it has been hypothesized that they would originate during spermatogenic mitotic divisions or during meiosis.

Doubtlessly, further research is needed to identify whether there are specific environmental or paternal host factors that are associated with paternally transmissible chromosomal abnormalities. Another fascinating challenge is posed by the lack of knowledge about whether there exist specific types of chromosomal abnormalities that are produced at a specific stage of germ cell production, the relative contribution that spermatogenetic mechanisms might exert on the development of chromosomal aberrations, and how both processes are affected by age.

Sperm DNA Damage

What makes DNA damage an extensively investigated topic is the irreplaceable nature of the DNA molecule. Vital information about cellular content and function is sheltered in the DNA, rendering it a crucial target for age-related degeneration. For instance, damage in the DNA can cause cell cycle arrest, cell death, or mutations the accumulation of which may lead to deregulation of transcription pathways, reduced fitness, and, ultimately, the aging phenotype. In spermatozoa, DNA damage could show up in the form of DNA fragmentation, abnormal chromatin packaging, and protamine deficiency potentially leading to cell impairment, and a number of studies have contributed to our understanding of whether an association with male aging exists. Higher levels of double-stranded DNA breaks were reported in older men (Singh et al. 2003), and a gradual age-related upward trend has been proposed for DNA fragmentation (Wyrobek et al. 2006) since the DNA fragmentation index (DFI) more than doubled in men between 20 and 60 years old. As for DNA fragmentation, data in the literature are not completely homogeneous concerning its relationship with paternal age, but there is undeniable ever-increasing evidence for a DFI augment with advancing age (Belloc et al. 2009; Plastira et al. 2007; Schmid et al. 2012; Vagnini et al. 2007).

In the myriad of DNA changes that occur as a consequence of aging, several theories collocate oxidative stress among those mechanisms predicted to play a causal role. Similarly to somatic cells, the continuous generation of reactive oxygen species (ROSs) produces oxidative damage, especially in spermatozoa, because of their high content of polyunsaturated fatty acid in the cell membrane (Aitken and Krausz 2001). Since ROSs production is likely to increase with age, it is plausible to hypothesize that, in men of advanced age, growing oxidative stress might be responsible for the age-related augment in sperm DNA damage. Moreover, changes in the efficiency of mismatch repair, base excision repair, nucleotide excision repair, and double-strand break repair mechanisms might endure the effect of aging and present themselves as cofactors in age-inflicted DNA damage. In conclusion, paternal age does indeed appear to be an additional factor that is positively correlated with an increase in DNA damage in spermatozoa deriving from men of both fertile

and infertile couples (Sartorius and Nieschlag 2010). Clearly, further research should be conducted to better define not only the nature but also the mechanisms underlying age-related changes in DNA and the extent of the damage that could be consequently produced.

Effect of Father's Age on Disease Risk

It is now fully recognized that children born from older parents are exposed to a much higher risk of having genetic disorders. This has been extensively proven for women delivering children at advanced ages, as witnessed by the strong maternal age effect for Down syndrome. However, there is an ever-growing evidence that paternal age also confers to offspring a susceptibility to a broad range of conditions, including spontaneous dominant disorders, congenital anomalies, neurological diseases (i.e., schizophrenia and autism), and some types of childhood cancers.

De Novo Mutation Rate in Male Gametes

Paternal aging is considered the major cause of new mutations in human populations (Crow 1999). Indeed, it is common knowledge that male germ cells undergo continuous cell divisions, which clearly accumulate with age, consequently leading to an accelerated mutation rate in spermatozoa. This could be due to several mechanisms, the first of which are the alterations of age-sensitive processes such as DNA replication and repair (Crow 2000). Moreover, the accumulation of mutagens from either external or internal sources, which would certainly increase with age, might also contribute to the increased occurrence of DNA replication inaccuracy.

Information available on de novo mutation rates mainly derives from studies in which the direct examination of parent-to-child transmission is limited to testing specific genes or regions, whereas the innovative whole-genome, sequencing-based studies are still inadequate to address this question. A recent study by Kong et al. addressed this issue by performing an estimate of the genome-wide mutation rate by sequencing the entire genomes of 78 parent-offspring trios (Kong et al. 2012). In particular, they focused on single nucleotide polymorphisms (SNPs), showing that the transmission of mutations to children is mainly due to fathers, and this behavior seemed closely linked to the paternal age. Considering that in this study the father's average age was 29.7 years old, the mean of the de novo mutation rate of SNPs was 1.20×10^{-8} per nucleotide per generation (Kong et al. 2012). This effect increased with the father's age (approximately two mutations per year), and the risk that children carrying harmful mutations, which could potentially lead to pathological conditions, increased proportionally.

Although in some circumstances the evidence for an association with advanced paternal age is not always consistent and reproducible, this is not the case for a small group of conditions known as paternal age effect (PAE) disorders, of which Apert syndrome and achondroplasia are considered the best representative examples. PAE disorders include some other disorders due to specific mutations in the fibroblastic growth factor receptor (*FGFR*) genes: mutations in *FGFR2* cause Apert, Crouzon, and Pfeiffer syndromes, mutations in *FGFR3* cause achondroplasia, thanatophoric dysplasia, hypochondroplasia, and Muenke syndrome. All these conditions are caused by substitution: transition/transversion at CpG dinucleotides or transition/transversion at non-CpG dinucleotides at key points within the growth factor receptor-RAS signaling pathway. These syndromes are characterized by autosomal dominant transmission; 1:30,000/130,000 birth prevalence for new mutations; paternal origin of mutations; strong paternal age effect; and a high apparent germ line mutation rate. Apert syndrome is a form of acrocephalosyndactyly, characterized by malformations of the skull, face, hands, and feet. In most cases there are two different mutations in the germ line occurring in the *FGFR2* gene: C to G at position 755, and C to G at position 758, which cause, respectively, a serine to tryptophan and a proline to arginine change in the protein (Wilkie et al. 1995). Achondroplasia is a common cause of dwarfism, and more than 99 % of the cases are caused by two different mutations in the *FGFR3* gene. In about 98 % of the cases, a G to A point mutation at nucleotide 1138 of the *FGFR3* gene causes a glycine-to-arginine substitution (Rousseau et al. 1996) and a G to C point mutation at nucleotide 1138 causes about 1 % of cases. These point mutations originate from unaffected fathers, suggesting that these events take place in the spermatogonial stem cells during spermatogenesis. The common explanation for these effects is the copy-error hypothesis, which postulates an accumulation of recurrent mutations in specific DNA hotspots. Although this process may play a specific role, alone it cannot explain these paternal age effects (Goriely and Wilkie 2012). Using a new polymerase chain reaction approach, it was possible to reveal that, although the mutational events in Apert syndrome are rare, when they take place, they become enriched because the encoded mutant proteins confer a selective advantage on spermatogonial cells, originating the so-called protein-driven *selfish* selection (Goriely et al. 2005). This mechanism is better known for somatic mutations that occur during neoplasia rather than in germ line diseases. In fact, if the same mutations that take place in PAE disorders occurred in somatic cells, they would lead to neoplasia: 755C>G and 758C>G substitutions in the *FGFR2* gene define endometrial cancer, and 1138G>A and 1138G>C in the *FGFR3* gene cause bladder cancer (Goriely and Wilkie 2012). Thus, it is important to consider that these mutational events may lead to both a specific syndrome and an oncogenetic process.

In conclusion, what are the long-term consequences of selfish selection? It seems that with age, spermatozoa of all men are progressively enriched with PAE mutations, even though PAE disorders fortunately have a low reproductive fitness; conversely, other mutations that define mild syndromes can be transmitted over many generations, representing a contribution to genetic variability.

Telomeres: The Bright Side of Aging

Telomeres are specific DNA sequences enclosing a number of (TTAGGG)_n repeats located at the ends of all chromosomes. Although their function is not yet fully established, it is widely known that telomeres are involved in the protection of chromosomes from fusion, recombination, and degradation. In many tissues, telomere length (TL) is shortened by successive cell divisions, and consequently it tends to progressively diminish with age. Therefore, TL changes are believed to be implicated in cell senescence and aging as well as tumorigenesis and DNA repair (De Meyer et al. 2007; Unryn et al. 2005). Consistent with this, elderly people, whose leukocytes display shorter telomeres due to their advanced age, are presumably subjected to a higher morbidity and reduced life expectancy.

Although it is well known that TL reduces with age in most proliferating tissues, spermatozoa represent the exception to the rule. For instance, there is substantial evidence that sperm TL dynamics follows a fascinating divergent trajectory that entails the elongation of sperm telomeres with age (Aston et al. 2012; Baird et al. 2006; Kimura et al. 2008). This notion provides a completely novel facet of the effects that might be exerted by paternal age and demonstrates that sometimes clouds do have a silver lining. In fact, emerging data provide growing evidence that older fathers will transmit to their offspring longer leukocyte telomeres (Arbeev et al. 2011; De Meyer et al. 2007; Unryn et al. 2005). In addition, a recent study performed on delayed human reproduction found that such an association between paternal age and offspring's TL is cumulative across multiple generations since the paternal grandfather's age predicts longer telomeres in grandchildren at their father's birth ($p=0.038$) (Eisenberg et al. 2012). The most common explanation for telomere lengthening among offspring of older fathers is the high telomerase activity in the testes (Baird et al. 2006; Kimura et al. 2008). Aston et al. (2012) suggest that sperm TL elongation is dependent on an overactivation of telomerase in male germ cells, leading to TL lengthening at every replication cycle (estimated value = 2.48 bp/replication). However, it remains to be defined why testicular telomerase would lead to the progressive lengthening of sperm telomeres rather than just maintain a stable length. Kimura et al. (2008) proposed that testicular telomerase exerts a preferential effect on long telomeres. This might depend on the fact that spermatozoa displaying short telomeres undergo a negative selection that with age progressively leads to their disproportional extinction (Kimura et al. 2008).

In light of these data, telomere lengthening might be considered the bright side of aging because delayed fatherhood would not only imply negative consequences, but could also confer positive traits to future generations conceived by aged fathers such as higher survival and lower risk of developing TL-related diseases.

Conclusions

Human aging includes a number of time-related processes occurring throughout adult life that guide a wide range of physiological changes that increase an individual's vulnerability to death and weaken normal functions and intensify one's susceptibility to a number of diseases. The ever-spreading phenomenon of postponing parenthood till older ages represents one of the multiple aspects of the aging process, given the recognition of advanced parental age effects.

While extensive evidence has proven maternal age to be a major and independent negative factor for fertility, the effects of paternal age remain poorly understood. However, there is growing evidence that advanced paternal age correlates with a number of complications, and 40 years old has been proposed as the “amber light” in a man's reproductive life. Reproductive studies suggest that male aging does not affect a couple's fecundity as an independent factor but that its effects manifest themselves in combination with maternal age or in the presence of altered spermatogenesis. This information might depend on the difficulty in discriminating paternal from maternal age effects, implying the need to direct further research straight to the male germ cells. This sperm-specific approach helped to define a pure paternal age effect on a multitude of issues discussed throughout this chapter (Fig. 2.1). Current data in the literature suggest that the spermatozoa of aged men apparently more frequently

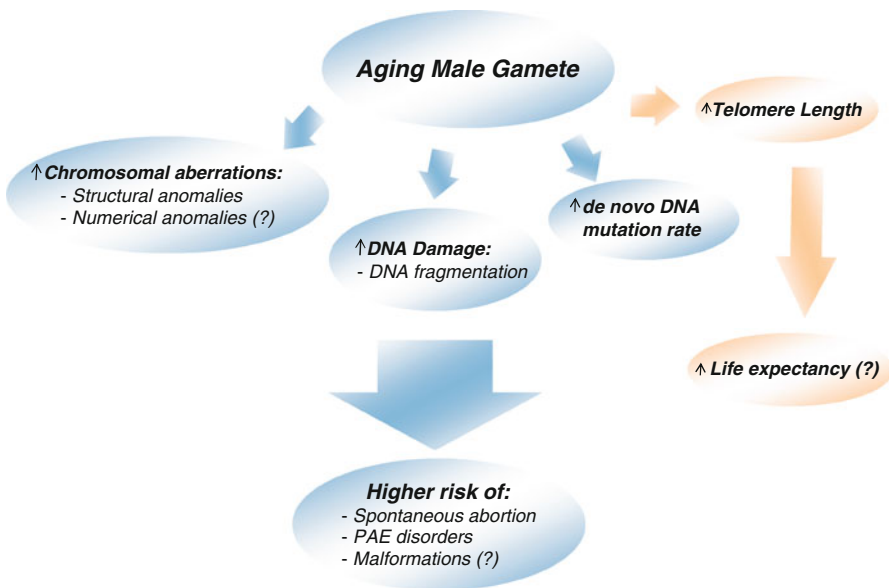


Fig. 2.1 Graphical summary of various consequences reportedly derived from the process of aging in the male gamete

undergo age-related modifications, potentially leading to various consequences. The occurrence of such alterations in male germ cells has rather important implications because any damage to reproductive cells might produce permanent effects not only on the fertility status of the questioned patient but also on the health and viability of the offspring, with potential consequences on the fitness of future generations.

References

- Aitken RJ, Krausz C (2001) Oxidative stress, DNA damage and the Y chromosome. *Reproduction* 122(4):497–506, Cambridge
- Allen JW, Collins BW, Setzer RW (1996) Spermatid micronucleus analysis of aging effects in hamsters. *Mutat Res* 316(5–6):261–266
- Arbeev KG, Hunt SC, Kimura M et al (2011) Leukocyte telomere length, breast cancer risk in the offspring: the relations with father's age at birth. *Mech Ageing Dev* 132(4):149–153
- Asada H, Sueoka K, Hashiba T et al (2000) The effects of age and abnormal sperm count on the nondisjunction of spermatozoa. *J Assist Reprod Genet* 17(1):51–59
- Aston KI, Hunt SC, Susser E et al (2012) Divergence of sperm and leukocyte age-dependent telomere dynamics: implications for male-driven evolution of telomere length in humans. *Mol Hum Reprod* 18(11):517–522
- Baird DM, Britt-Compton B, Rowson J et al (2006) Telomere instability in the male germline. *Hum Mol Genet* 15(1):45–51
- Baptista J, Mercer C, Prigmore E et al (2008) Breakpoint mapping and array CGH in translocations: comparison of a phenotypically normal and an abnormal cohort. *Am J Hum Genet* 82(4):927–936
- Belloc S, Benkhalifa M, Junca AM et al (2009) Paternal age and sperm DNA decay: discrepancy between chromomycin and aniline blue staining. *Reprod Biomed Online* 19(2):264–269
- Bosch M, Rajmil O, Martínez-Pasarell O et al (2001) Linear increase of diploidy in human sperm with age: a four-colour FISH study. *Eur J Hum Genet* 9(7):533–538
- Bosch M, Rajmil O, Egozcue J et al (2003) Linear increase of structural and numerical chromosome 9 abnormalities in human sperm regarding age. *Eur J Hum Genet* 11(10):754–759
- Brandriff B, Gordon LA, Crawford BB et al (1988) Sperm chromosome analysis to assess potential germ cell mosaicism. *Clin Genet* 34(2):85–89
- Crow JF (1999) Spontaneous mutation in man. *Mutat Res* 437(1):5–9
- Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* 1(1):40–47
- de la Rochebrochard E, Thonneau P (2002) Paternal age and maternal age are risk factors for miscarriage; results of a multicentre European study. *Hum Reprod* 17(6):1649–1656
- De La Rochebrochard E, McElreavey K, Thonneau P (2003) Paternal age over 40 years: the “amber light” in the reproductive life of men? *J Androl* 24(4):459–465
- De Meyer T, Rietzschel ER, De Buyzere ML et al (2007) Paternal age at birth is an important determinant of offspring telomere length. *Hum Mol Genet* 16(24):3097–3102
- Eisenberg DT, Hayes MG, Kuzawa CW (2012) Delayed paternal age of reproduction in humans is associated with longer telomeres across two generations of descendants. *Proc Natl Acad Sci USA* 109(26):10251–10256
- Estop AM, Márquez C, Munné S et al (1995) An analysis of human sperm chromosome breakpoints. *Am J Hum Genet* 56(2):452–460
- Fisch H (2009) Older men are having children, but the reality of a male biological clock makes this trend worrisome. *Geriatrics* 64(1):14–17
- Goriely A, Wilkie AO (2012) Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am J Hum Genet* 90(2):175–200

- Goriely A, McVean GA, van Pelt AM et al (2005) Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia. *Proc Natl Acad Sci USA* 102(17):6051–6056
- Griffin DK, Abruzzo MA, Millie EA et al (1995) Non-disjunction in human sperm: evidence for an effect of increasing paternal age. *Hum Mol Genet* 4(12):2227–2232
- Guttenbach M, Köhn FM, Engel W et al (2000) Meiotic nondisjunction of chromosomes 1, 17, 18, X, and Y in men more than 80 years of age. *Biol Reprod* 63(6):1727–1729
- Hassold TJ (1998) Nondisjunction in the human male. *Curr Top Dev Biol* 37:383–406
- Hassold THP (2001) To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet* 2(4):280–291
- Kimura M, Cherkas LF, Kato BS et al (2008) Offspring's leukocyte telomere length, paternal age, and telomere elongation in sperm. *PLoS Genet* 4(2):e37
- Kinakin B, Rademaker A, Martin R (1997) Paternal age effect of YY aneuploidy in human sperm, as assessed by fluorescence in situ hybridization. *Cytogenet Cell Genet* 78(2):116–119
- Kong A, Frigge ML, Masson G et al (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488(7412):471–475
- Kurahashi H, Bolor H, Kato T et al (2009) Recent advance in our understanding of the molecular nature of chromosomal abnormalities. *J Hum Genet* 54(5):253–260
- Lowe X, Collins B, Allen J et al (1995) Aneuploidies and micronuclei in the germ cells of male mice of advanced age. *Mutat Res* 338(1–6):59–76
- Lowe X, Eskenazi B, Nelson DO et al (2001) Frequency of XY sperm increases with age in fathers of boys with Klinefelter syndrome. *Am J Hum Genet* 69(5):1046–1054
- Martin RH, Rademaker AW (1987) The effect of age on the frequency of sperm chromosomal abnormalities in normal men. *Am J Hum Genet* 41(3):484–492
- Martin RH, Spriggs E, Ko E et al (1995) The relationship between paternal age, sex ratios, and aneuploidy frequencies in human sperm, as assessed by multicolor FISH. *Am J Hum Genet* 57(6):1395–1399
- Martin JA, Hamilton BE, Sutton PD et al (2007) Births: final data for 2005. *Natl Vital Stat Rep* 56(6):1–103
- McInnes B, Rademaker A, Greene CA et al (1998) Abnormalities for chromosomes 13 and 21 detected in spermatozoa from infertile men. *Hum Reprod* 13(10):2787–2790
- Ohye T, Inagaki H, Kogo H et al (2010) Paternal origin of the de novo constitutional t(11;22)(q23;q11). *Eur J Hum Genet* 18(7):783–787
- Olson SD, Megenis RE (1988) Preferential parental origin of de novo structural chromosomal rearrangements. In: *The cytogenetics of mammalian autosomal rearrangements*. Alan R. Liss, New York, pp 583–599
- Pacchierotti F, Andreozzi U, Russo A et al (1983) Reciprocal translocations in ageing mice and in mice with long-term low-level ²³⁹Pu contamination. *Int J Radiat Biol Relat Stud Phys Chem Med* 43(4):445–450
- Plastira K, Msaouel P, Angelopoulou R et al (2007) The effects of age on DNA fragmentation, chromatin packaging and conventional semen parameters in spermatozoa of oligoasthenoteratozoospermic patients. *J Assist Reprod Genet* 24(10):437–443
- Robbins WA, Vine MF, Truong KY et al (1997) Use of fluorescence in situ hybridization (FISH) to assess effects of smoking, caffeine, and alcohol on aneuploidy load in sperm of healthy men. *Environ Mol Mutagen* 30(2):175–183
- Rousseau F, Bonaventure J, Legeai-Mallet L et al (1996) Mutations of the fibroblast growth factor receptor-3 gene in achondroplasia. *Horm Res* 45(1–2):108–110
- Rousseaux S, Hazzouri M, Pelletier R et al (1998) Disomy rates for chromosomes 14 and 21 studied by fluorescent in-situ hybridization in spermatozoa from three men over 60 years of age. *Mol Hum Reprod* 4(7):695–699
- Rubes J, Lowe X, Moore D 2nd et al (1998) Smoking cigarettes is associated with increased sperm disomy in teenage men. *Fertil Steril* 70(4):715–723
- Rudak E, Jacobs PA, Yanagimachi R (1978) Direct analysis of the chromosome constitution of human spermatozoa. *Nature* 274(5674):911–913

- Sartorelli EM, Mazzucatto LF, de Pina-Neto JM (2001) Effect of paternal age on human sperm chromosomes. *Fertil Steril* 76(6):1119–1123
- Sartorius GA, Nieschlag E (2010) Paternal age and reproduction. *Hum Reprod Update* 16(1):65–79
- Schmid TE, Grant PG, Marchetti F et al (2012) Elemental composition of human semen is associated with motility and genomic sperm defects among older men. *Hum Reprod* 28:274–282, Oxford, England
- Seymour F, Duffy C, Korner A (1935) A case of authentic fertility in a man of 94. *JAMA* 105:1423–1424
- Shi Q, Martin RH (2000) Aneuploidy in human sperm: a review of the frequency and distribution of aneuploidy, effects of donor age and lifestyle factors. *Cytogenet Cell Genet* 90(3–4):219–226
- Singh NP, Muller CH, Berger RE (2003) Effects of age on DNA double-strand breaks and apoptosis in human sperm. *Fertil Steril* 80(6):1420–1430
- Sloter E, Nath J, Eskenazi B et al (2004) Effects of male age on the frequencies of germinal and heritable chromosomal abnormalities in humans and rodents. *Fertil Steril* 81(4):925–943
- Sloter ED, Marchetti F, Eskenazi B et al (2007) Frequency of human sperm carrying structural aberrations of chromosome 1 increases with advancing age. *Fertil Steril* 87(5):1077–1086
- Tarin JJ, Brines J, Cano A (1998) Long-term effects of delayed parenthood. *Hum Reprod* 13(9):2371–2376
- Templado C, Donate A, Giraldo J et al (2011) Advanced age increases chromosome structural abnormalities in human spermatozoa. *Eur J Hum Genet* 19(2):145–151
- Thomas NS, Durkie M, Van Zyl B et al (2006) Parental and chromosomal origin of unbalanced de novo structural chromosome abnormalities in man. *Hum Genet* 119(4):444–450
- Thomas NS, Morris JK, Baptista J et al (2012) De novo apparently balanced translocations in man are predominantly paternal in origin and associated with a significant increase in paternal age. *J Med Genet* 47(2):112–115
- Tiemann-Boege I, Navidi W, Grewal R et al (2002) The observed human sperm mutation frequency cannot explain the achondroplasia paternal age effect. *Proc Natl Acad Sci USA* 99(23):14952–14957
- Unryn BM, Cook LS, Riabowol KT (2005) Paternal age is positively linked to telomere length of children. *Aging Cell* 4(2):97–101
- Vagnini L, Baruffi RL, Mauri AL et al (2007) The effects of male age on sperm DNA damage in an infertile population. *Reprod Biomed Online* 15(5):514–519
- Wiener-Megnazi Z, Auslender R, Dirnfeld M (2012) Advanced paternal age and reproductive outcome. *Asian J Androl* 14(1):69–76
- Wilkie AO, Slaney SF, Oldridge M et al (1995) Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome. *Nat Genet* 9(2):165–172
- Wyrobek AJ, Eskenazi B, Young S et al (2006) Advancing age has differential effects on DNA damage, chromatin integrity, gene mutations, and aneuploidies in sperm. *Proc Natl Acad Sci USA* 103(25):9601–9606